PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/82, 15/60, 9/88, 1/21, C12Q 1/68, A01H 5/00 — A2

(11) International Publication Number: WO 00/00627

(43) International Publication Date: 6 January 2000 (06.01.00)

(21) International Application Number: PCT/US99/14777

(22) International Filing Date: 25 June 1999 (25.06.99)

(30) Priority Data:
60/090,924 26 June 1998 (26.06.98) US
60/121,965 26 February 1999 (26.02.99) US

(71) Applicant *(for all designated States except US)*: CALGENE LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: MATSUI, Kenji [JP/JP]; 2–51, Nishiki–cho, Yamaguchi City, Yamaguchi 7653–0068 (JP).

(74) Agent: SCHWEDLER, Carl, J.; 1920 Fifth Street, Davis, CA 95616 (US).

(81) Designated States: CA, CN, JP, KR, MX, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: FATTY ACID HYDROPEROXIDE LYASE NUCLEIC ACID SEQUENCES

(57) Abstract

This invention relates to plant Hydroperoxide Lyase or HPO lyase polynucleotides and polypeptides. DNA constructs useful for the expression of a plant HPO lyase in a cell are described. Furthermore, DNA constructs useful for the antisense expression of a plant HPO lyase in a plant cell are described. Such constructs will contain a DNA sequence encoding the plant HPO lyase of interest under the control of regulatory elements capable of preferentially directing the expression of the plant HPO lyase in plant tissue, when such a construct is expressed in a transgenic plant. This invention also relates to methods of using a DNA sequence encoding a plant HPO lyase for the modification of the volatile aldehydes in plant tissues, as well as for methods of increasing disease resistance in a plant.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Fatty Acid Hydroperoxide Lyase Nucleic Acid Sequences

INTRODUCTION

This application claims the benefit of U.S. Provisional Application Number 60/090,924 filed June 26, 1998, U.S. Provisional Application Number 60/121,965 filed February 26, 1999, and U.S. Provisional Application Number 60/121,968 filed February 26, 1999.

Technical Field

This invention relates to the application of genetic engineering techniques to plants. More specifically, the invention relates to plant hydroperoxide lyase sequences and methods for the use of such sequences.

Background

With the development of genetic engineering techniques, it is now possible to transfer genes from a variety of organism into the genome of a large number of different plant species. This process has many advantages over plant breeding techniques, as genes may now be transferred from one plant species to another plant species, rather than simply from a plant to the same, or different, but closely related, species.

Degradation of polyunsaturated fatty acids starts by the oxygenation atcis-cis double bonds of polyunsaturated fatty acids. This reaction is catalyzed by lipoxygenase (EC 1.13.11.12) enzymes which are present in plants, animals and microorganisms. The oxygenated products, called fatty acid hydroperoxides, are precursors for many important hormones (e.g. lipoxins, jasmonic acid, traumatic acid) and flavor/fragrance molecules (e.g. cis-3-hexenol, 1-octen-3-ol) in plants.

Compounds, such as jasmonic acid, are produced from hydroperoxides, such as 13-hydroperoxylinolenic acid, via an allene oxide synthase (referred to as AOS) and an allene oxide cyclase (referred to as ACS)-dependent pathway. Jasmonic acid is involved in stress and disease resistance signaling responses via the octadecanoid pathway. 13-hydroperoxylinolenic can also be catabolized by peroxygenases to form cutin monomers.

Alternatively, 13-hydroperoxylinolenic can be catabolyzed by hydroperoxide lyase eventually forming volatile aldehydes and traumatic acid.

Fatty acid hydroperoxide lyase (HPO lyase) catalyzes the cleavage of carbon-carbon bonds in polyunsaturated fatty acid hyroperoxides to produce short-chain aldehydes and ω-oxo-acids (Vick, *et al.* (1976) *Plant Physiol.* 57:780-788). The products of lysis of fatty acid hydroperoxides, such as short-chain volatile aldehydes are common in plant species. The short-chain volatile aldehydes contribute to the "green notes" in a wide variety of plant leaves, vegetables and fruits. "Green notes" are volatile molecules that contribute to the organoleptic qualities of flavor and fragrance of edible plant tissues. These qualities are often referred to as grassy, or "green" characteristics. Other short-chain volatile aldehydes, such as (3Z, 6Z)-nonadienol produced by the lysis of fatty acid 9-hydroperoxide by a fatty acid 9-hydroperoxide lyase (9-HPO lyase or 9-HPOL), contribute a melon aroma and/or a melon flavor, or sometimes referred to as a "melon" or "fresh" characteristic, to fruits and vegetables. Such characteristics are important to industries concerned with fragrances and flavorings.

Furthermore, short-chain aldehydes are also thought to be involved in disease resistance. For example, Croft, *et al* ((1993) *Plant Physiol.* 101:13-24) recently reported that (3*Z*)-hexenol and (2*E*)-hexenal levels increased during a hypersensitive-response in kidney bean plants. In addition, they also demonstrated that (2*E*)-hexenal is an effective antibacterial agent.

The characterization of hydroperoxide (also referred to as HPO lyase or HPOL) is useful for the further study of plant fatty acid metabolism systems and for the development of transgenic plant with increased organoleptic properties, including aromas and flavors. Studies of plant mechanisms may provide means to further enhance, control, modify, or otherwise alter the organoleptic qualities of edible plant tissues. Furthermore, the elucidation of the physiological roles of HPO lyase and it's products may be useful for the further study of disease resistance responses, such as the HR response. Of particular interest are the nucleic acid sequences of genes encoding proteins which may be useful for applications in genetic engineering.

SUMMARY OF THE INVENTION

The present invention is directed to hydroperoxide lyase (also referred to herein as HPO lyase and HPOL), and in particular to HPO lyase polynucleotides. The polynucleotides of the present invention include those derived from plant sources.

In one aspect of the present invention, polynucloetides are provided which encode HPO lyase polypeptides. In particular, polynucleotides are provided which encode 13-HPO lyase polypeptides, and polynucleotides are provided encoding 9-HPO lyase polypeptides.

One aspect of the present invention relates to oligonucleotides which include partial or complete HPO lyase encoding sequences.

It is also an aspect of the present invention to provide recombinant DNA constructs which can be used for transcription or transcription and translation (expression) of HPO lyase. In particular, constructs are provided which are capable of transcription or transcription and translation in host cells. Particularly preferred constructs are those capable of transcription or transcription and translation in plant cells.

In another aspect of the present invention, methods are provided for production of HPO lyase in a host cell or progeny thereof. In particular, host cells are transformed or transfected with a DNA construct which can be used for transcription or transcription and translation of HPO lyase. The recombinant cells which contain HPO lyase are also part of the present invention. Particularly preferred host cells include yeast, bacterial, insect, and plant cells.

In a further aspect, the present invention relates to methods of using polynucleotide and polypeptide sequences to modify the volatile content of host cells. Preferred host cells of the present invention include bacterial, yeast, insect and plant host cells. Host cells having such a modified volatile content are also contemplated herein.

In yet a further aspect of the present invention, methods of using polynucleotide and polypeptide sequences of the present invention to produce host plants having an altered response to diseases are provided.

The modified host cells and oils obtained by the expression or suppression of the HPO lyase proteins are also considered part of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows the complete nucleotide sequence of the *Arabidopsis* HPO lyase.

Figure 2 shows a comparison of the amino acid sequences of the bell pepper HPO lyase and the *Arabidopsis* HPO lyase-like sequence.

Figure 3 shows a comparison of the amino acid sequences of the *Arabidopsis* allene oxide synthase and the *Arabidopsis* HPO lyase-like sequence.

Figure 4 shows the complete nucleotide sequence of the tomato HPO lyase.

Figure 5 shows the complete nucleotide sequence of the cucumber allene oxide synthase.

Figure 6 shows the complete nucleotide sequence of the cucumber 9-Hydroperoxide Lyase.

Figure 7 shows the amino acid sequence alignment between the bell pepper, banana, and *Arabidopsis* HPO lyase, with the highly conserved peptide sequences highlighted.

Figure 8 Provides the percent similarity in the upper right corner and the percent divergence in the lower right corner for the nucleotide sequences (Figure 8A) and amino acid sequences (Figure 8B) of the bell pepper HPOL (CaHPOL), tomato fruit HPOL (LeHPOL), cucumber hypocotyl HPOL (CsC17HPOL, pseudogene), Arabidopsis inflorecence HPOL (AtHPOL), banana leaf HPOL (MsHPOL), cucumber hypocotyl 9-HPOL (Cs15HPOL), Guayule AOS (GuAOS), flaxseed AOS (LiAOS), and the Arabidopsis AOS (AtAOS).

Figure 9 shows the gas chromatography (GC) analysis of the cucumber 9-HPO lyase using linoleic acid 13-hydroperoxide (Figure 9A) and linoleic acid 9- hydroperoxide (figure 9B) substrates.

Figure 10 provides the results of the spectrophotometric assay of the cucumber 9-HPO lyase expressed from *E. coli* using linoleic acid 13-hydroperoxide and linoleic acid 9-hydroperoxide substrates.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to hydroperoxide lyase (also referred to herein as HPO lyase and HPOL), particularly the isolated HPO lyase nucleic acid sequences encoding the HPO lyase protein from plant sources. A hydroperoxide lyase of this invention includes any nucleic acid sequence encoding amino acids from a host cell source, such as a polypeptide, obtainable from a cell source, which demonstrates the ability to form short-chain aldehydes and oxo-acids from fatty acid hydroperoxides under plant enzyme reactive conditions. By

"enzyme reactive conditions" is meant that any necessary conditions are available in an environment (i.e., such factors as temperature, pH, lack of inhibiting substances) which will permit the enzyme to function. By "enzyme reactive conditions" is meant that any necessary conditions are available in an environment (i.e., such factors as temperature, pH, lack of inhibiting substances) which will permit the enzyme to function.

The term HPO lyase encoding sequence as applied herein refers to any polynucleotide sequence which encodes a hydroperoxide lyase polypeptide which is capable of producing short-chain aldehydes and oxo-acids from fatty acid hydroperoxides. HPO lyase encoding sequence can encode polypeptides having preferential activity towards particular fatty acid hydroperoxides. Particular fatty acid hydroperoxides include, but are not limited to fatty acid 13-hydroperoxide, and fatty acid 9-hydroperoxide.

As used herein, the term 13-hydroperoxide lyase (13-HPO lyase or 13-HPOL) refers to any enzyme which forms short chain aldehydes and oxo-acids from fatty acid 13-hydroperoxides. Examples of short chain aldehydes include, but are not limited to cis 3-Hexenal and examples of oxo-acids include, but are not limited to 12-oxo-(9Z)dodecenoic acid. An example of fatty acid 13-hydroperoxide include, but is not limited to Linolenic acid 13-hydroperoxide.

As used herein, the term 9-hydroperoxide lyase (9-HPO lyase or 9-HPOL) refers to any enzyme which forms short chain aldehydes and oxo-acids from fatty acid 9-hydroperoxides. Examples of short chain aldehydes include, but are not limited to (3Z,6Z)-nonadienal and examples of oxo-acids include, but are not limited to 9-oxo-nonanoic acid. An example of fatty acid 9-hydroperoxide include, but is not limited to Linolenic acid 9-hydroperoxide or 9-hydroperoxy-(10E, 12Z, 15Z)-octadecadienoic acid.

Isolated proteins, Polypeptides and Polynucleotides

A first aspect of the present invention relates to isolated HPO lyase polynucleotides. The polynucleotide sequences of the present invention include isolated polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from the group of sequences set forth in the Sequence Listing and to other polynucleotide sequences closely related to such sequences and variants thereof.

The invention provides a polynucleotide sequence identical over its entire length to each coding sequence as set forth in the Sequence Listing. The invention also provides the coding sequence for the mature polypeptide or a fragment thereof, as well as the coding

sequence for the mature polypeptide or a fragment thereof in a reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, pro-, or prepro- protein sequence. The polynucleotide can also include non-coding sequences, including for example, but not limited to, non-coding 5' and 3' sequences, such as the transcribed, untranslated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence that encodes additional amino acids. For example, a marker sequence can be included to facilitate the purification of the fused polypeptide. Polynucleotides of the present invention also include polynucleotides comprising a structural gene and the naturally associated sequences that control gene expression.

The invention also includes polynucleotides of the formula:

$$X\text{-}(R_1)_n\text{-}(R_2)\text{-}(R_3)_n\text{-}Y$$

wherein, at the 5' end, X is hydrogen, and at the 3' end, Y is hydrogen or a metal, R_1 and R_3 are any nucleic acid residue, n is an integer between 1 and 3000, preferably between 1 and 1000 and R_2 is a nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from the group set forth in the Sequence Listing. In the formula, R_2 is oriented so that its 5' end residue is at the left, bound to R_1, and its 3' end residue is at the right, bound to R_3. Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

The invention also relates to variants of the polynucleotides described herein that encode for variants of the polypeptides of the invention. Variants that are fragments of the polynucleotides of the invention can be used to synthesize full-length polynucleotides of the invention. Preferred embodiments are polynucleotides encoding polypeptide variants wherein 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues of a polypeptide sequence of the invention are substituted, added or deleted, in any combination. Particularly preferred are substitutions, additions, and deletions that are silent such that they do not alter the properties or activities of the polynucleotide or polypeptide.

Further preferred embodiments of the invention that are at least 50%, 60%, or 70% identical over their entire length to a polynucleotide encoding a polypeptide of the invention, and polynucleotides that are complementary to such polynucleotides. More preferable are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the invention and polynucleotides that are complementary thereto. In this regard, polynucleotides at least 90% identical over their entire

length are particularly preferred, those at least 95% identical are especially preferred. Further, those with at least 97% identity are highly preferred and those with at least 98% and 99% identity are particularly highly preferred, with those at least 99% being the most highly preferred.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptides encoded by the polynucleotides set forth in the Sequence Listing.

The invention further relates to polynucleotides that hybridize to the above-described sequences. In particular, the invention relates to polynucleotides that hybridize under stringent conditions to the above-described polynucleotides. As used herein, the terms "stringent conditions" and "stringent hybridization conditions" mean that hybridization will generally occur if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/milliliter denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at approximately 65°C. Other hybridization and wash conditions are well known and are exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, cold Spring Harbor, NY (1989), particularly Chapter 11.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set for in the Sequence Listing under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or a fragment thereof; and isolating said polynucleotide sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers as described herein.

As discussed herein regarding polynucleotide assays of the invention, for example, polynucleotides of the invention can be used as a hybridization probe for RNA, cDNA, or genomic DNA to isolate full length cDNAs or genomic clones encoding a polypeptide and to isolate cDNA or genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in the Sequence Listing. Such probes will generally comprise at least 15 bases. Preferably such probes will have at least 30 bases and can have at least 50 bases. Particularly preferred probes will have between 30 bases and 50 bases, inclusive.

The coding region of each gene that comprises or is comprised by a polynucleotide sequence set forth in the Sequence Listing may be isolated by screening using a DNA sequence provided in the Sequence Listing to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to identify members of the library which hybridize to the probe. For example, synthetic oligonucleotides are prepared which correspond to the HPO lyase EST sequences. The oligonucleotides are used as primers in polymerase chain reaction (PCR) techniques to obtain 5' and 3' terminal sequence of HPO lyase genes. Alternatively, where oligonucleotides of low degeneracy can be prepared from particular HPO lyase peptides, such probes may be used directly to screen gene libraries for HPO lyase gene sequences. In particular, screening of cDNA libraries in phage vectors is useful in such methods due to lower levels of background hybridization.

Typically, a HPO lyase sequence obtainable from the use of nucleic acid probes will show 60-70% sequence identity between the target HPO lyase sequence and the encoding sequence used as a probe. However, lengthy sequences with as little as 50-60% sequence identity may also be obtained. The nucleic acid probes may be a lengthy fragment of the nucleic acid sequence, or may also be a shorter, oligonucleotide probe. When longer nucleic acid fragments are employed as probes (greater than about 100 bp), one may screen at lower stringencies in order to obtain sequences from the target sample which have 20-50% deviation (i.e., 50-80% sequence homology) from the sequences used as probe. Oligonucleotide probes can be considerably shorter than the entire nucleic acid sequence encoding an HPO lyase enzyme, but should be at least about 10, preferably at least about 15, and more preferably at least about 20 nucleotides. A higher degree of sequence identity is desired when shorter regions are used as opposed to longer regions. It may thus be desirable to identify regions of highly conserved amino acid sequence to design oligonucleotide probes for detecting and recovering other related HPO lyase genes. Shorter probes are often particularly useful for polymerase chain reactions (PCR), especially when highly conserved sequences can be identified. (*See*, Gould, *et al.*, *PNAS USA* (1989) *86*:1934-1938.).

Another aspect of the present invention relates to HPO lyase polypeptides. Such polypeptides include isolated polypeptides set forth in the Sequence Listing, as well as polypeptides and fragments thereof, particularly those polypeptides which exhibit HPO lyase activity and also those polypeptides which have at least 50%, 60% or 70% identity, preferably at least 80% identity, more preferably at least 90% identity, and most preferably at least 95%

identity to a polypeptide sequence selected from the group of sequences set forth in the Sequence Listing, and also include portions of such polypeptides, wherein such portion of the polypeptide preferably includes at least 30 amino acids and more preferably includes at least 50 amino acids.

"Identity", as is well understood in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods including, but not limited to, those described in *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M. and Griffin, H.G., eds., Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press (1987); *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., Stockton Press, New York (1991); and Carillo, H., and Lipman, D., SIAM *J Applied Math*, 48:1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available programs. Computer programs which can be used to determine identity between two sequences include, but are not limited to, GCG (Devereux, J., et al., *Nucleic Acids Research* 12(1):387 (1984); suite of five BLAST programs, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN) (Coulson, *Trends in Biotechnology, 12:* 76-80 (1994); Birren, *et al., Genome Analysis, 1:* 543-559 (1997)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., *et al.*, NCBI NLM NIH, Bethesda, MD 20894; Altschul, S., *et al., J. Mol. Biol.*, 215:403-410 (1990)). The well known Smith Waterman algorithm can also be used to determine identity.

Parameters for polypeptide sequence comparison typically include the following:

Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48:443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, *Proc. Natl. Acad. Sci USA* 89:10915-10919 (1992)

Gap Penalty: 12

Gap Length Penalty: 4

A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters along with no penalty for end gap are the default parameters for peptide comparisons.

Parameters for polynucleotide sequence comparison include the following:

Algorithm: Needleman and Wunsch, J. Mol. Biol. 48:443-453 (1970)

Comparison matrix: matches = +10; mismatches = 0

Gap Penalty: 50

Gap Length Penalty: 3

A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters are the default parameters for nucleic acid comparisons.

The invention also includes polypeptides of the formula:

$$X\text{-}(R_1)_n\text{-}(R_2)\text{-}(R_3)_n\text{-}Y$$

wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R_1 and R_3 are any amino acid residue, n is an integer between 1 and 1000, and R_2 is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in the Sequence Listing. In the formula, R_2 is oriented so that its amino terminal residue is at the left, bound to R_1, and its carboxy terminal residue is at the right, bound to R_3. Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

The polypeptides of the present invention can be mature protein or can be part of a fusion protein.

Fragments and variants of the polypeptides are also considered to be a part of the invention. A fragment is a variant polypeptide which has an amino acid sequence that is entirely the same as part but not all of the amino acid sequence of the previously described polypeptides. The fragments can be "free-standing" or comprised within a larger polypeptide of which the fragment forms a part or a region, most preferably as a single continuous region. Preferred fragments are biologically active fragments which are those fragments that mediate activities of the polypeptides of the invention, including those with similar activity or improved activity or with a decreased activity. Also included are those fragments that antigenic or immunogenic in an animal, particularly a human.

Variants of the polypeptide also include polypeptides that vary from the sequences set forth in the Sequence Listing by conservative amino acid substitutions, substitution of a residue by another with like characteristics. In general, such substitutions are among Ala, Val, Leu and Ile; between Ser and Thr; between Asp and Glu; between Asn and Gln; between Lys and Arg; or between Phe and Tyr. Particularly preferred are variants in which 5 to 10; 1 to 5; 1 to 3 or one amino acid(s) are substituted, deleted, or added, in any combination.

Variants that are fragments of the polypeptides of the invention can be used to produce the corresponding full length polypeptide by peptide synthesis. Therefore, these variants can be used as intermediates for producing the full-length polypeptides of the invention.

The polynucleotides and polypeptides of the invention can be used, for example, in the transformation of host cells, such as plant host cells, as further discussed herein.

The invention also provides polynucleotides that encode a polypeptide that is a mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids within the mature polypeptide (for example, when the mature form of the protein has more than one polypeptide chain). Such sequences can, for example, play a role in the processing of a protein from a precursor to a mature form, allow protein transport, shorten or lengthen protein half-life, or facilitate manipulation of the protein in assays or production. It is contemplated that cellular enzymes can be used to remove any additional amino acids from the mature protein.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. The inactive precursors generally are activated when the prosequences are removed. Some or all of the prosequences may be removed prior to activation. Such precursor protein are generally called proproteins.

In the examples provided below, a nucleic acid sequence from *Arabidopsis* is identified from Genbank which is highly homologous to a bell pepper HPO lyase. The *Arabidopsis* sequence had been previously reported to encode allene oxide synthase.

Nucleic acid sequence comparisons between the bell pepper HPO lyase, an *Arabidopsis* allene oxide synthase (Laudert, *et al.* (1996) *Plant Mol Biol* 31(2)323-335), and the *Arabidopsis* sequence from Genbank indicates that the *Arabidopsis* sequence is more similar to the bell pepper HPO lyase than to the allene oxide synthase.

Also provided in the examples below, a nucleic acid sequence from cucumber (*Cucumis sativus*) is identified from cDNA libraries made from total RNA isolated from

cucumber hypocotyls. A full length coding sequence is obtained, and the product encoded by the full length sequence demonstrates activity towards the substrate linolenic acid 9-hydroperoxide to produce (3Z, 6Z)-nonadienal and 9-oxo-nonanoic acid.

Plant Constructs and Methods of Use

Of interest is the use of the nucleotide sequences in recombinant DNA constructs to direct the transcription or transcription and translation (expression) of the HPO lyase sequences of the present invention in a host cell. Of particular interest is the use of the polynucleotide sequences of the present invention in recombinant DNA constructs to direct the transcription or transcription and translation (expression) of the HPO lyase sequences of the present invention in a host plant cell.

The expression constructs generally comprise a promoter functional in a host cell operably linked to a nucleic acid sequence encoding an HPO lyase of the present invention and a transcriptional termination region functional in a host plant cell. Of particular interest is the use of promoters (also referred to as transcriptional initiation regions) functional in plant host cells.

Those skilled in the art will recognize that there are a number of promoters which are functional in plant cells, and have been described in the literature. Chloroplast and plastid specific promoters, chloroplast or plastid functional promoters, and chloroplast or plastid operable promoters are also envisioned.

One set of promoters functional in plant cells are constitutive promoters such as the CaMV35S or FMV35S promoters that yield high levels of expression in most plant organs. Enhanced or duplicated versions of the CaMV35S and FMV35S promoters are useful in the practice of this invention (Odell, *et al.* (1985) *Nature* 313:810-812; Rogers, U.S. Patent Number 5,378, 619). In addition, it may also be preferred to bring about expression of the HPO lyase gene in specific tissues of the plant, such as leaf, stem, root, tuber, seed, fruit, etc., and the promoter chosen should have the desired tissue and developmental specificity.

Of particular interest is the expression of the nucleic acid sequences of the present invention from transcription initiation regions which are preferentially expressed in a plant seed tissue. Examples of such seed preferential transcription initiation sequences include those sequences derived from sequences encoding plant storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Examples of such promoters include the 5'

regulatory regions from such genes as napin (Kridl *et al.*, *Seed Sci. Res. 1*:209:219 (1991)), phaseolin, zein, soybean trypsin inhibitor, ACP, stearoyl-ACP desaturase, soybean α' subunit of β-conglycinin (soy 7s, (Chen *et al.*, *Proc. Natl. Acad. Sci.*, 83:8560-8564 (1986))) and oleosin.

It may be advantageous to direct the localization of proteins conferring HPO lyase to a particular subcellular compartment, for example, to the mitochondrion, endoplasmic reticulum, vacuoles, chloroplast or other plastidic compartment. For example, where the genes of interest of the present invention will be targeted to plastids, such as chloroplasts, for expression, the constructs will also employ the use of sequences to direct the gene to the plastid. Such sequences are referred to herein as chloroplast transit peptides (CTP) or plastid transit peptides (PTP). In this manner, where the gene of interest is not directly inserted into the plastid, the expression construct will additionally contain a gene encoding a transit peptide to direct the gene of interest to the plastid. The chloroplast transit peptides may be derived from the gene of interest, or may be derived from a heterologous sequence having a CTP. Such transit peptides are known in the art. See, for example, Von Heijne *et al.* (1991) *Plant Mol. Biol. Rep. 9*:104-126; Clark *et al.* (1989) *J. Biol. Chem. 264*:17544-17550; della-Cioppa *et al.* (1987) *Plant Physiol. 84*:965-968; Romer *et al.* (1993) *Biochem. Biophys. Res Commun. 196*:1414-1421; and, Shah *et al.* (1986) *Science 233*:478-481.

Depending upon the intended use, the constructs may contain the nucleic acid sequence which encodes the entire HPO lyase protein, or a portion thereof. For example, where antisense inhibition of a given HPO lyase protein is desired, the entire HPO lyase sequence is not required. Furthermore, where HPO lyase sequences used in constructs are intended for use as probes, it may be advantageous to prepare constructs containing only a particular portion of a HPO lyase encoding sequence, for example a sequence which is discovered to encode a highly conserved HPO lyase region.

The skilled artisan will recognize that there are various methods for the inhibition of expression of endogenous sequences in a host cell. Such methods include, but are not limited to antisense suppression (Smith, *et al.* (1988) *Nature* 334:724-726) , co-suppression (Napoli, *et al.* (1989) *Plant Cell* 2:279-289), ribozymes (PCT Publication WO 97/10328), and combinations of sense and antisense Waterhouse, *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964. Methods for the suppression of endogenous sequences in a host cell typically employ the transcription or transcription and translation of at least a portion of the

sequence to be suppressed. Such sequences may be homologous to coding as well as non-coding regions of the endogenous sequence.

Regulatory transcript termination regions may be provided in plant expression constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding the HPO lyase or a convenient transcription termination region derived from a different gene source, for example, the transcript termination region which is naturally associated with the transcript initiation region. The skilled artisan will recognize that any convenient transcript termination region which is capable of terminating transcription in a plant cell may be employed in the constructs of the present invention.

Alternatively, constructs may be prepared to direct the expression of the HPO lyase sequences directly from the host plant cell plastid. Such constructs and methods are known in the art and are generally described, for example, in Svab, *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:8526-8530 and Svab and Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917 and in U.S. Patent Number 5,693,507.

A plant cell, tissue, organ, or plant into which the recombinant DNA constructs containing the expression constructs have been introduced is considered transformed, transfected, or transgenic. A transgenic or transformed cell or plant also includes progeny of the cell or plant and progeny produced from a breeding program employing such a transgenic plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of a HPO lyase nucleic acid sequence.

Plant expression or transcription constructs having an HPO lyase as the DNA sequence of interest for increased or decreased expression thereof may be employed with a wide variety of plant life, particularly, plant life involved in the production of vegetable oils for edible and industrial uses. Most especially preferred are seed, fruit, vegetable and leaf crops. Plants of interest include, but are not limited to, rapeseed (Canola and High Erucic Acid varieties), sunflower, safflower, cotton, soybean, peanut, coconut and oil palms, corn tomato, strawberry, bell pepper and melon.. Depending on the method for introducing the recombinant constructs into the host cell, other DNA sequences may be required. Importantly, this invention is applicable to dicotyledyons and monocotyledons species alike and will be readily applicable to new and/or improved transformation and regulation techniques.

Of particular interest, is the use of HPO lyase constructs in plants to produce plants or plant parts, including, but not limited to leaves, stems, roots, reproductive, and seed, with a modified volatilecontent.

It is contemplated that the gene sequences may be synthesized, either completely or in part, especially where it is desirable to provide plant-preferred sequences. Thus, all or a portion of the desired structural gene (that portion of the gene which encodes the HPO lyase protein) may be synthesized using codons preferred by a selected host. Host-preferred codons may be determined, for example, from the codons used most frequently in the proteins expressed in a desired host species.

One skilled in the art will readily recognize that antibody preparations, nucleic acid probes (DNA and RNA) and the like may be prepared and used to screen and recover "homologous" or "related" sequences from a variety of plant sources. Homologous sequences are found when there is an identity of sequence, which may be determined upon comparison of sequence information, nucleic acid or amino acid, or through hybridization reactions between a known HPO LYASE and a candidate source. Conservative changes, such as Glu/Asp, Val/Ile, Ser/Thr, Arg/Lys and Gln/Asn may also be considered in determining sequence homology. Amino acid sequences are considered homologous by as little as 25% sequence identity between the two complete mature proteins. (*See* generally, Doolittle, R.F., *OF URFS and ORFS* (University Science Books, CA, 1986.)

Thus, other HPO lyase may be obtained from the specific sequences provided herein. Furthermore, it will be apparent that one can obtain natural and synthetic sequences, including modified amino acid sequences and starting materials for synthetic-protein modeling from the exemplified HPO lyase sequences and from sequences which are obtained through the use of such exemplified sequences. Modified amino acid sequences include sequences which have been mutated, truncated, increased and the like, whether such sequences were partially or wholly synthesized. Sequences which are actually purified from plant preparations or are identical or encode identical proteins thereto, regardless of the method used to obtain the protein or sequence, are equally considered naturally derived.

For immunological screening, antibodies to the protein can be prepared by injecting rabbits or mice with the purified protein or portion thereof, such methods of preparing antibodies being well known to those in the art. Either monoclonal or polyclonal antibodies can be produced, although typically polyclonal antibodies are more useful for gene isolation. Western analysis may be conducted to determine that a related protein is present in a crude

extract of the desired plant species, as determined by cross-reaction with the antibodies to the encoded proteins. When cross-reactivity is observed, genes encoding the related proteins are isolated by screening expression libraries representing the desired plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*, Second Edition (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

Potential host cells include both prokaryotic and eukaryotic cells. A host cell may be unicellular or found in a multicellar differentiated or undifferentiated organism depending upon the intended use. Cells of this invention may be distinguished by having an HPO lyase foreign to the wild-type cell present therein, for example, by having a recombinant nucleic acid construct encoding a HPO lyase therein not native to the host species. Preferred host cells include bacterial, yeast, insect, mammalian, and plant host cells. Particularly preferred host cells include yeast and plant host cells.

Prokaryotic cells include gram negative as well as gram positive bacteria, for example E. coli, and B. subtilis strains. Suitable examples are well known to the skilled artisan. As described in more detail in the examples that follow, an HPO lyase isolated from cucumber hypocotyl is expressed in E coli, strain M15. The protein expressed from the E coli is capable of producing the aldehyde 3(Z)-nonenal and 2(E)-nonenal from linoleic acid 9-hydroperoxide. Thus, the HPO lyase isolated from the cucumber hypocotyl encodes a 9-HPO lyase.

Eukaryotic host cells include fungi, including yeasts, insect cells, and plant cells. Methods for the expression of DNA sequences of interest in yeast cells are known in the art and are generally described in "Guide to yeast genetics and molecular biology",Guthrie and Fink, eds. Methods in enzymology , Academic Press, Inc. Vol 194 (1991) and Gene expression technology", Goeddel ed, Methods in Enzymology, Academic Press, Inc., Vol 185 (1991). In addition, methods for the expression of HPO lyase genes are described in European patent Application EP 0 801 133 A2, the entirety of which is incorporated herein by reference.

The fungal recombinant vector may be any vector which can be conveniently subjected to recombinant DNA procedures. The choice of a vector will typically depend on the compatibility of the vector with the fungal host cell into which the vector is to be introduced. The vector may be a linear or a closed circular plasmid. The vector system may be a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the fungal host.

The fungal vector may be an autonomously replicating vector, *i.e.*, a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the fungal cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. For integration, the vector may rely on the nucleic acid sequence of the vector for stable integration of the vector into the genome by homologous or nonhomologous recombination. Alternatively, the vector may contain additional nucleic acid sequences for directing integration by homologous recombination into the genome of the fungal host. The additional nucleic acid sequences enable the vector to be integrated into the host cell genome at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, there should be preferably two nucleic acid sequences which individually contain a sufficient number of nucleic acids, preferably 400bp to 1500bp, more preferably 800bp to 1000bp, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. These nucleic acid sequences may be any sequence that is homologous with a target sequence in the genome of the fungal host cell and, furthermore, may be non-encoding or encoding sequences.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. Examples of origin of replications for use in a yeast host cell are the 2 micron origin of replication and the combination of CEN3 and ARS 1. Any origin of replication may be used which is compatible with the fungal host cell of choice.

The fungal vectors of the present invention preferably contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides, for example biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs and the like. The selectable marker may be selected from the group including, but not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferase), *bar* (phosphinothricin acetyltransferase), hygB (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphatedecarboxylase) and *sC* (sulfate adenyltransferase) and *trpC* (anthranilate synthase). Preferred for use in an *Aspergillus* cell are the *amdS* and *pyrG* markers of *Aspergillus nidulans* or *Aspergillus oryzae* and the bar marker of *Streptomyces hygroscopicus*. Furthermore, selection may be

accomplished by co-transformation, *e.g.*, as described in WO 91/17243, the entirety of which is herein incorporated by reference. A nucleic acid sequence of the present invention may be operably linked to a suitable promoter sequence. The promoter sequence is a nucleic acid sequence which is recognized by the fungal host cell for expression of the nucleic acid sequence. The promoter sequence contains transcription and translation control sequences which mediate the expression of the protein or fragment thereof.

A promoter may be any nucleic acid sequence which shows transcriptional activity in the fungal host cell of choice and may be obtained from genes encoding polypeptides either homologous or heterologous to the host cell. Examples of suitable promoters for directing the transcription of a nucleic acid construct of the invention in a filamentous fungal host are promoters obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase and hybrids thereof. In a yeast host, a useful promoter is the *Saccharomyces cerevisiae* enolase (eno-1) promoter. Particularly preferred promoters are the TAKA amylase, NA2-tpi (a hybrid of the promoters from the genes encoding *Aspergillus niger* neutral alpha -amylase and *Aspergillus oryzae* triose phosphate isomerase) and glaA promoters.

A protein or fragment thereof encoding nucleic acid molecule of the present invention may also be operably linked to a terminator sequence at its 3' terminus. The terminator sequence may be native to the nucleic acid sequence encoding the protein or fragment thereof or may be obtained from foreign sources. Any terminator which is functional in the fungal host cell of choice may be used in the present invention, but particularly preferred terminators are obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase and *Saccharomyces cerevisiae* enolase.

A protein or fragment thereof encoding nucleic acid molecule of the present invention may also be operably linked to a suitable leader sequence. A leader sequence is a nontranslated region of a mRNA which is important for translation by the fungal host. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the protein or fragment thereof. The leader sequence may be native to the nucleic acid sequence encoding the protein or fragment thereof or may be obtained from foreign sources.

Any leader sequence which is functional in the fungal host cell of choice may be used in the present invention, but particularly preferred leaders are obtained from the genes encoding *Aspergillus oryzae* TAKA amylase and *Aspergillus oryzae* triose phosphate isomerase.

A polyadenylation sequence may also be operably linked to the 3' terminus of the nucleic acid sequence of the present invention. The polyadenylation sequence is a sequence which when transcribed is recognized by the fungal host to add polyadenosine residues to transcribed mRNA. The polyadenylation sequence may be native to the nucleic acid sequence encoding the protein or fragment thereof or may be obtained from foreign sources. Any polyadenylation sequence which is functional in the fungal host of choice may be used in the present invention, but particularly preferred polyadenylation sequences are obtained from the genes encoding *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase and *Aspergillus niger* alpha-glucosidase.

To avoid the necessity of disrupting the cell to obtain the protein or fragment thereof and to minimize the amount of possible degradation of the expressed protein or fragment thereof within the cell, it is preferred that expression of the protein or fragment thereof gives rise to a product secreted outside the cell. To this end, a protein or fragment thereof of the present invention may be linked to a signal peptide linked to the amino terminus of the protein or fragment thereof. A signal peptide is an amino acid sequence which permits the secretion of the protein or fragment thereof from the fungal host into the culture medium. The signal peptide may be native to the protein or fragment thereof of the invention or may be obtained from foreign sources. The 5' end of the coding sequence of the nucleic acid sequence of the present invention may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region which encodes the secreted protein or fragment thereof. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region which is foreign to that portion of the coding sequence which encodes the secreted protein or fragment thereof. The foreign signal peptide may be required where the coding sequence does not normally contain a signal peptide coding region. Alternatively, the foreign signal peptide may simply replace the natural signal peptide to obtain enhanced secretion of the desired protein or fragment thereof. The foreign signal peptide coding region may be obtained from a glucoamylase or an amylase gene from an *Aspergillus* species, a lipase or proteinase gene from *Rhizomucor miehei*, the gene for the alpha-factor from *Saccharomyces cerevisiae*, or the calf preprochymosin gene. An effective signal peptide for fungal host cells is the *Aspergillus oryzae* TAKA amylase

signal, *Aspergillus niger* neutral amylase signal, the *Rhizomucor miehei* aspartic proteinase signal, the *Humicola lanuginosus* cellulase signal, or the *Rhizomucor miehei* lipase signal. However, any signal peptide capable of permitting secretion of the protein or fragment thereof in a fungal host of choice may be used in the present invention.

A protein or fragment thereof encoding nucleic acid molecule of the present invention may also be linked to a propeptide coding region. A propeptide is an amino acid sequence found at the amino terminus of aproprotein or proenzyme. Cleavage of the propeptide from the proprotein yields a mature biochemically active protein. The resulting polypeptide is known as a propolypeptide or proenzyme (or a zymogen in some cases). Propolypeptides are generally inactive and can be converted to mature active polypeptides by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide or proenzyme. The propeptide coding region may be native to the protein or fragment thereof or may be obtained from foreign sources. The foreign propeptide coding region may be obtained from the *Saccharomyces cerevisiae* alpha-factor gene or *Myceliophthora thermophila* laccase gene (WO 95/33836, the entirety of which is herein incorporated by reference).

The procedures used to ligate the elements described above to construct the recombinant expression vector of the present invention are well known to one skilled in the art (see, for example, Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor, N.Y., (1989)).

The present invention also relates to recombinant fungal host cells produced by the methods of the present invention which are advantageously used with the recombinant vector of the present invention. The cell is preferably transformed with a vector comprising a nucleic acid sequence of the invention followed by integration of the vector into the host chromosome. The choice of fungal host cells will to a large extent depend upon the gene encoding the protein or fragment thereof and its source. The fungal host cell may, for example, be a yeast cell or a filamentous fungal cell.

"Yeast" as used herein includes *Ascosporogenous* yeast (*Endomycetales*), *Basidiosporogenous* yeast and yeast belonging to the *Fungi Imperfecti* (*Blastomycetes*). The *Ascosporogenous* yeasts are divided into the families *Spermophthoraceae* and *Saccharomycetaceae*. The latter is comprised of four subfamilies, *Schizosaccharomycoideae* (for example, genus *Schizosaccharomyces*), *Nadsonioideae, Lipomycoideae* and *Saccharomycoideae* (for example, genera *Pichia, Kluyveromyces* and *Saccharomyces*). The *Basidiosporogenous* yeasts include the genera *Leucosporidim, Rhodosporidium,*

Sporidiobolus, Filobasidium and *Filobasidiella*. Yeast belonging to the *Fungi Imperfecti* are divided into two families, *Sporobolomycetaceae* (for example, genera *Sorobolomyces* and *Bullera*) and *Cryptococcaceae* (for example, genus *Candida*). Since the classification of yeast may change in the future, for the purposes of this invention, yeast shall be defined as described in Biology and Activities of Yeast (Skinner *et al.*, *Soc. App. Bacteriol. Symposium Series* No. 9, (1980), the entirety of which is herein incorporated by reference). The biology of yeast and manipulation of yeast genetics are well known in the art (*see,* for example, *Biochemistry and Genetics of Yeast,* Bacil *et al.* (ed.), 2nd edition, 1987; *The Yeasts*, Rose and Harrison (eds.), 2nd ed., (1987); and *The Molecular Biology of the Yeast Saccharomyces*, Strathern *et al.* (eds.), (1981), all of which are herein incorporated by reference in their entirety).

The recombinant fungal host cells of the present invention may further comprise one or more sequences which encode one or more factors that are advantageous in the expression of the protein or fragment thereof, for example, an activator (e.g., a trans-acting factor), a chaperone and a processing protease. The nucleic acids encoding one or more of these factors are preferably not operably linked to the nucleic acid encoding the protein or fragment thereof. An activator is a protein which activates transcription of a nucleic acid sequence encoding a polypeptide (Kudla *et al., EMBO* 9:1355-1364(1990); Jarai and Buxton, *Current Genetics* 26:2238-244(1994); Verdier, *Yeast 6*:271-297(1990), all of which are herein incorporated by reference in their entirety). The nucleic acid sequence encoding an activator may be obtained from the genes encoding *Saccharomyces cerevisiae* heme activator protein 1 (hap1), *Saccharomyces cerevisiae* galactose metabolizing protein 4 (gal4) and *Aspergillus nidulans* ammonia regulation protein (areA). For further examples, see Verdier, *Yeast 6*:271-297 (1990); MacKenzie *et al.*, *Journal of Gen. Microbiol. 139*:2295-2307 (1993), both of which are herein incorporated by reference in their entirety). A chaperone is a protein which assists another protein in folding properly (Hartl *et al., TIBS 19*:20-25 (1994); Bergeron *et al., TIBS* 19:124-128 (1994); Demolder *et al., J. Biotechnology* 32:179-189 (1994); Craig, *Science 260*:1902-1903(1993); Gething and Sambrook, *Nature* 355:33-45 (1992); Puig and Gilbert, *J Biol. Chem. 269*:7764-7771 (1994); Wang and Tsou, *FASEB Journal* 7:1515-11157 (1993); Robinson *et al.*, *Bio/Technology 1*:381-384 (1994), all of which are herein incorporated by reference in their entirety). The nucleic acid sequence encoding a chaperone may be obtained from the genes encoding *Aspergillus oryzae* protein disulphide isomerase, *Saccharomyces cerevisiae* calnexin, *Saccharomyces cerevisiae* BiP/GRP78 and

Saccharomyces cerevisiae Hsp70. For further examples, *see* Gething and Sambrook, *Nature 355*:33-45 (1992); Hartl *et al., TIBS 19*:20-25 (1994). A processing protease is a protease that cleaves a propeptide to generate a mature biochemically active polypeptide (Enderlin and Ogrydziak, *Yeast 10*:67-79 (1994); Fuller *et al., Proc. Natl. Acad. Sci. (U.S.A.) 86*:1434-1438 (1989); Julius *et al., Cell 37*:1075-1089 (1984); Julius *et al., Cell 32*:839-852 (1983), all of which are incorporated by reference in their entirety). The nucleic acid sequence encoding a processing protease may be obtained from the genes encoding *Aspergillus niger* Kex2, *Saccharomyces cerevisiae* dipeptidylaminopeptidase, *Saccharomyces cerevisiae* Kex2 and *Yarrowia lipolytica* dibasic processing endoprotease (xpr6). Any factor that is functional in the fungal host cell of choice may be used in the present invention.

Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts and regeneration of the cell wall in a manner known per se. Suitable procedures for transformation of *Aspergillus* host cells are described in EP 238 023 and Yelton *et al., Proc. Natl. Acad. Sci. (U.S.A.) 81*:1470-1474 (1984), both of which are herein incorporated by reference in their entirety. A suitable method of transforming *Fusarium* species is described by Malardier *et al., Gene 78:*147-156 (1989), the entirety of which is herein incorporated by reference. Yeast may be transformed using the procedures described by Becker and Guarente, In: Abelson and Simon, (eds.), *Guide to Yeast Genetics and Molecular Biology, Methods Enzymol.* Volume 194, pp 182-187, Academic Press, Inc., New York; Ito *et al., J. Bacteriology 153*:163 (1983); Hinnen *et al., Proc. Natl. Acad. Sci. (U.S.A.) 75*:1920 (1978), all of which are herein incorporated by reference in their entirety.

The present invention also relates to methods of producing the protein or fragment thereof comprising culturing the recombinant fungal host cells under conditions conducive for expression of the protein or fragment thereof. The fungal cells of the present invention are cultivated in a nutrient medium suitable for production of the protein or fragment thereof using methods known in the art. For example, the cell may be cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing the protein or fragment thereof to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art (*see, e.g.*, Bennett and LaSure (eds.), *More Gene Manipulations in Fungi*, Academic Press, CA, (1991), the entirety of which is herein incorporated by reference). Suitable media are available from commercial

suppliers or may be prepared according to published compositions (*e.g.*, in catalogues of the American Type Culture Collection, Manassas, VA). If the protein or fragment thereof is secreted into the nutrient medium, a protein or fragment thereof can be recovered directly from the medium. If the protein or fragment thereof is not secreted, it is recovered from cell lysates.

The expressed protein or fragment thereof may be detected using methods known in the art that are specific for the particular protein or fragment. These detection methods may include the use of specific antibodies, formation of an enzyme product, or disappearance of an enzyme substrate. For example, if the protein or fragment thereof has enzymatic activity, an enzyme assay may be used. Alternatively, if polyclonal or monoclonal antibodies specific to the protein or fragment thereof are available, immunoassays may be employed using the antibodies to the protein or fragment thereof. The techniques of enzyme assay and immunoassay are well known to those skilled in the art.

The resulting protein or fragment thereof may be recovered by methods known in the arts. For example, the protein or fragment thereof may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation. The recovered protein or fragment thereof may then be further purified by a variety of chromatographic procedures, e.g., ion exchange chromatography, gel filtration chromatography, affinity chromatography, or the like.

Methods for the expression of DNA sequences of interest in insect host cells are also well known in the art, and are reviewed by Lucow and Summers, (1988) *Bio/technology* 6:47-55, the entirety of which is incorporated herein by reference.

To confirm the activity and specificity of the proteins encoded by the identified nucleic acid sequences as HPO lyase enzymes, *in vitro* assays can be performed in insect cell cultures using baculovirus expression systems. Such baculovirus expression systems are known in the art and are described by Lee, *et al.* U.S. Patent Number 5,348,886, the entirety of which is herein incorporated by reference.

In addition, other expression constructs may be prepared to assay for protein activity utilizing different expression systems. Such expression constructs are transformed into yeast or prokaryotic host and assayed for HPO lyase activity. Such expression systems are known in the art and are readily available through commercial sources.

The method of transformation in obtaining such transgenic plants is not critical to the instant invention, and various methods of plant transformation are currently available. Furthermore, as newer methods become available to transform crops, they may also be directly applied hereunder. For example, many plant species naturally susceptible to *Agrobacterium* infection may be successfully transformed via tripartite or binary vector methods of *Agrobacterium* mediated transformation. In many instances, it will be desirable to have the construct bordered on one or both sides by T-DNA, particularly having the left and right borders, more particularly the right border. This is particularly useful when the construct uses *A. tumefaciens* or *A. rhizogenes* as a mode for transformation, although the T-DNA borders may find use with other modes of transformation. In addition, techniques of microinjection, DNA particle bombardment, and electroporation have been developed which allow for the transformation of various monocot and dicot plant species.

Normally, included with the DNA construct will be a structural gene having the necessary regulatory regions for expression in a host and providing for selection of transformant cells. The gene may provide for resistance to a cytotoxic agent, e.g. antibiotic, heavy metal, toxin, etc., complementation providing prototrophy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species the expression construct or components thereof are introduced, one or more markers may be employed, where different conditions for selection are used for the different hosts.

Where *Agrobacterium* is used for plant cell transformation, a vector may be used which may be introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium* host. The Ti- or Ri-plasmid containing the T-DNA for recombination may be armed (capable of causing gall formation) or disarmed (incapable of causing gall formation), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cells and gall.

In some instances where *Agrobacterium* is used as the vehicle for transforming host plant cells, the expression or transcription construct bordered by the T-DNA border region(s) will be inserted into a broad host range vector capable of replication in *E. coli* and *Agrobacterium*, there being broad host range vectors described in the literature. Commonly used is pRK2 or derivatives thereof. See, for example, Ditta, *et al.*, (*Proc. Nat. Acad. Sci., U.S.A.* (1980) *77*:7347-7351) and EPA 0 120 515, which are incorporated herein by reference. Alternatively, one may insert the sequences to be expressed in plant cells into a vector

containing separate replication sequences, one of which stabilizes the vector in *E. coli*, and the other in *Agrobacterium*. See, for example, McBride and Summerfelt (*Plant Mol. Biol.* (1990) *14*:269-276), wherein the pRiHRI (Jouanin, *et al., Mol. Gen. Genet.* (1985) 201:370-374) origin of replication is utilized and provides for added stability of the plant expression vectors in host *Agrobacterium* cells.

Included with the expression construct and the T-DNA will be one or more markers, which allow for selection of transformed Agrobacterium and transformed plant cells. A number of markers have been developed for use with plant cells, such as resistance to chloramphenicol, kanamycin, the aminoglycoside G418, hygromycin, or the like. The particular marker employed is not essential to this invention, one or another marker being preferred depending on the particular host and the manner of construction.

For transformation of plant cells using *Agrobacterium*, explants may be combined and incubated with the transformed *Agrobacterium* for sufficient time for transformation, the bacteria killed, and the plant cells cultured in an appropriate selective medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of vegetable oils.

Thus, in another aspect of the present invention, methods for modifying the volatile composition of a host cell. In general the methods involve either increasing or decreasing the levels of volatile compounds in a host cell. The method generally comprises the use of expression constructs to direct the expression of the polynucleotides of the present invention in a host cell.

Of particular interest is the use of expression constructs to modify the levels of volatile compounds in a host plant cell. Most particular, the methods find use in modifying the levels of volatile compounds in plant parts including, but not limited to, leaves, roots, stems, flowers, tuber, fruits, legumes, seeds, and seed oils obtained from plant seeds.

In another embodiment of the present invention, expression constructs are provided which direct the expression of nucleic acid sequences encoding HPO lyase from *Arabidopsis* in bacterial and plant tissues.

Of particular interest in the present invention, is the use of such expression constructs to produce transgenic plants with increased production of short-chain volatile aldehydes in plant fruits and tissues. Such volatile aldehydes are important constituents of the

characteristic flavors of fruits, vegetables and green leaves (also referred to as "green notes"). Thus, the HPO lyase sequence of the present invention may be used in expression constructs to produce transgenic plants with improved green note flavor characteristics.

In order to increase lipid peroxidation, and thereby increasing "green note" and/or "melon" flavors/fragrances, in a plant tissue, coexpression of a plant or other 9-HPO lyase and/or 13-HPO lyase in a plant tissue with a second gene involved in lipid peroxidation may also find use in the present invention. For example, coexpression of a 13-HPO lyase and/or 9-HPO lyase sequence in a plant tissue with a DNA sequence encoding for another protein involved in lipid peroxidation, such as a lipoxygenase may increase lipid peroxidation and increase the total short-chain aldehydes produced in the plant tissue. Such an increase in short-chain aldehydes may increase the "green note" and/or "melon" flavor in an edible plant tissue.

Host cells expressing the 9-HPO lyases of the present invention provide a novel source of volatile aldehydes contributing to "green" and/or "melon" notes for use in various applications. Furthermore, the host cells may also contain constructs providing for a increased production of enzymes involved in lipid peroxidation, for example lipoxygenase. In addition, the host cells may also produce an increased amount of a particular fatty acid, or have a general increase in fatty acids. Such host cells may be obtained using traditional breeding techniques, including mutagenesis, as well as hosts genetically engineered with such an altered fatty acid composition.

Furthermore, plant host cells containing a construct providing for the expression of the HPO lyase sequences of the present invention find use as a source for aldehydes in reactions for the production of alcohols for use in flavorings and aromatic products. Such methods are known in the art and are described for example in U.S. Patent Number 5,695,973 and in PCT Publication WO 95/26413 the entireties of which are incorporated herein by reference. Generally, a mixture of aldehydes and alcohols are obtained from such methods. The methods generally involve a reaction mixture containing at least one unsaturated fatty acid, a plant material having a relatively high amount of enzyme activity of lipoxygenase and hydroperoxide lyase, and a source of alcohol dehydrogenase.

The unsaturated fatty acid may vary and include a single unsaturated fatty acid species as well as mixtures of several unsaturated fatty acids. The fatty acids are provided in a free acid form, and examples include, but are not limited to oleic acid, linoleic acid, linolenic acid (alpha and gamma forms), arachidonic acid, eicosapentaenoic acid, and ricinoleic acid.

Sources of the alcohol dehydrogenase include yeasts, as well as non-yeast molds. The alcohol dehydrogenase has the ability to convert an aldehyde to an alcohol. The yeast and non-yeast molds further provide a source of nicotine adenine dinucleotide (NADH) as a reducing agent.

The nucleic acid sequences of the present invention may also find use in expression constructs for the production of transgenic plants with increased resistance to various pathogens. Transgenic plants expressing the HPO lyase sequence of the present invention may exhibit an enhanced hypersensitive-reaction (HR response) in response to pathogen attack due to the increased production of aldehydes involved in the HR response, such as (3*Z*)-hexenal and (2*E*)-hexenal (Croft, *et al.* (1993) *Plant Physiol.* 101:13-24). Aldehydes, such as (2*E*)-hexenal, have also been shown to be effective anti-bacterial agents, further contributing to enhanced disease resistance (Croft, *et al.* (1993), *supra*). Furthermore, these compounds may be involved in a general wounding response in plants.

Also of particular interest in the present invention is the use of 9-HPO lyase nucleic acid sequences in constructs to direct the expression of 9-HPO lyase in a prokaryotic and/or eukaryotic host cells for the production of flavorings and aromas.

The invention now being generally described, it will be more readily understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

EXAMPLES

Example 1 Identification of Arabidopsis HPO Lyase Sequences

A nucleic acid fragment encoding hydroperoxide lyase from Bell pepper has been previously cloned and sequenced (Matsui, *et al.* (1996) *FEBS Letters* 394:21-24). The nucleotide sequence was used to search Genbank for HPO lyase related sequences. One accession identified from Genbank (Accession Number Z97339, (http://www.ncbi.nlm.nih.gov/web/Genbank/Index.html)) containing a genomic sequence from *Arabidopsis* was reported to encode an allene oxide synthase (Laudert, *et al.*(1996) *Plant Mol. Biol.* 31:323-335).

Sequence comparisons between the bell pepper HPO lyase, *Arabidopsis* allene oxide synthase (Laudert, *et al.* (1996) *supra*) and the *Arabidopsis* HPO lyase-like sequence from Genbank using Genetyx Mac (Software Development Co. Ltd.) indicated that the *Arabidopsis*

HPO lyase-like sequence is more similar to the bell pepper HPO lyase (57% identity) (see Figure 2) than to the allene oxide synthase sequence (39% identity) (see Figure 3).

Example 2 Construction of *Arabidopsis* cDNA libraries

Total RNA from seedling, inflorescence, and silique tissues of *Arabidopsis thaliana* is isolated for use in construction of complementary (cDNA) libraries. The procedure is an adaptation of the DNA isolation protocol of Webb and Knapp (D.M. Webb and S.J. Knapp, (1990) Plant Molec. Reporter, 8, 180-185). The following description assumes the use of 1g fresh weight of tissue. Frozen seed tissue is powdered by grinding under liquid nitrogen. The powder is added to 10ml REC buffer (50mM Tris-HCl, pH 9, 0.8M NaCl, 10mM EDTA, 0.5% w/v CTAB (cetyltrimethyl-ammonium bromide)) along with 0.2g insoluble polyvinylpolypyrrolidone, and ground at room temperature. The homogenate is centrifuged for 5 minutes at 12,000 x*g* to pellet insoluble material. The resulting supernatant fraction is extracted with chloroform, and the top phase is recovered.

The RNA is then precipitated by addition of 1 volume RecP (50mM Tris-HCL pH9, 10mM EDTA and 0.5% (w/v) CTAB) and collected by brief centrifugation as before. The RNA pellet is redissolved in 0.4 ml of 1M NaCl. The RNA pellet is redissolved in water and extracted with phenol/chloroform. Sufficient 3M potassium acetate (pH 5) is added to make the mixture 0.3M in acetate, followed by addition of two volumes of ethanol to precipitate the RNA. After washing with ethanol, this final RNA precipitate is dissolved in water and stored frozen.

Alternatively, total RNAs may be obtained using TRIzol reagent (BRL Life Technologies, Gaithersburg, MD) following the manufacturers protocol.

Complementary DNAs (cDNA) are obtained from the RNAs using the Marathon cDNA Amplification Kit (Clontech, Palo Alto, CA) following the manufacturers directions.

Example 3 Cloning of HPO Lyase Sequences

In order to characterize the protein encoded by the *Arabidopsis* cDNA GenBank sequence, the entire coding region corresponding to the *Arabidopsis* HPO lyase-like cDNA was obtained. (Figure 1) Synthetic oligo-nucleotide primers are designed to amplify the 5' and 3' ends from the HPO lyase-like sequence from RNA obtained in Example 2. Primers are designed according to the *Arabidopsis* HPO lyase-like sequence and are used in Rapid Amplification of cDNA Ends (RACE) reactions (Frohman *et al.* (1988) *Proc. Natl. Acad. Sci.*

USA 85:8998-9002). Amplification of flanking sequences from cDNA clones are performed using the Marathon cDNA Amplification kit (Clontech) according to the manufacturers protocol.

A pair of primers were designed to amplify the 5' and 3' regions from the *Arabidopsis* HPO lyase-like cDNA from the libraries described in example 2 above. These two primers, HPOL28 (for 3' RACE, 5'-CGGTTCCTCTGCGCCTCTCTCGCCGGCG-3') and HPOL21 (for 5' RACE, 5'-GCGGAACCGGAGGACTAAAACGCAGC-3') are used in PCR reactions with Adapter specific primers (AP1 5'-CCATCCTAATACGACTCACTATAGGGC-3') provided in the Marathon cDNA Amplification Kit. For amplification of the 5' region of the HPO lyase-like cDNA the primers AP1 and HPOL 21 were used, and for the amplification of the 3' region the AP1 primer was used in a reaction with the primer HPOL28. The cycle conditions used are: 94°C for 1 minute followed by 5 rounds of 94°C for 5 seconds, 72°C for 4 minutes, followed by 5 rounds of amplification using 94°C for 5 seconds, 70°C for 4 minutes, and finally 25 cycles of 94°C for 5 seconds, and 68°C for 4 minutes.

A single fragment of 1100 bp was obtained from the 3'RACE reaction with RNA obtained from the silique tissue described above. To confirm that the PCR product contained sequence corresponding to the HPO lyase-like sequence, a second round of PCR reactions using the same conditions described above was performed with the gel purified 1100 bp fragment. A reaction was performed with the primers HPOL13 (5'-CTTGGCGTAGTTCCTCAGCCTCTTG-3') and AP2 (5'-ACTCACTATAGGGCTCGAGCGGC-3') to amplify an approximately 1000 bp fragment as a confirmation of the HPO lyase-like sequence. The reamplified 1000 bp fragment was gel purified and cloned into the pCR2.1 TOPO vector (Invitrogen, Carlsbad, CA) to create the plasmid pCGN8094.

The 5' RACE reaction produced many non-specific fragments. A 1000 bp fragment was excised from the gel and cloned into the pCR2.1 TOPO(Invitrogen) cloning vector to create the plasmid pCGN8091.

Figure 1 discloses the complete nucleotide sequence of the *Arabidopsis* HPO lyase. The cDNA has a sequence of 1687 bp, and 47 and 137 bp of 5'- and 3'-noncoding regions, respectively. There is a stop codon in the 5'-noncoding region in-frame with the initiation codon. The longest ORF encodes a polypeptide of 492 amino acids with a calculated molecular mass of 54851 Da.

In order to determine the expression pattern of the *Arabidopsis* HPO lyase sequence, Northern blot analysis of total RNA isolated from various organs, as well as wounded, methyl jasmonate, and pathogen challenged leaves, of *Arabidopsis* is performed.

Total RNA is isolated from rosetta leaves, leaves on stems, stems, inflorescence, green buds, closed flowers, open flowers, siliques (5-10 mm), siliques (<5 mm), and 2, 3, 4, 5, and 6 day after germination seedlings using TRIzol reagent (Life Technologies, Gaithersburg, MD). The isolated RNA samples (10-20 μg) are separated on a formaldehyde-agarose gel and transferred to Hybond-N (Amersham). The transferred RNA is hybridized overnight with a probe corresponding to the 5' half of the *Arabidopsis* HPO lyase cDNA at 65°C in 6x SSC, 5x Denhardt's solution, 0.2% SDS, 20μg/ml salmon sperm DNA, 20 mM sodium phosphate buffer, pH 7.0. The hybridized membranes are washed once with 2x SSC, 0.1% SDS at 60°C for 20 minutes, and twice with 0.25x SSC, 0.1% SDS at 60° C for 20 minutes each. A 1.6 kb transcript corresponding to the HPO lyase gene is observed in RNA isolated from all the tissue sources examined. The highest level of expression is observed in the inflorescence. Additional bands of approximately 3.0 and 3.3 kb are observed possibly due to read-through of the intron-exon junction during transcription. In addition, Rojo, *et al.* (1998) *Plant Journal* 13:153-165 reports the hybridization of two mRNA bands of different size on Northern blots using a probe of HPO lyase.

To determine if the HPO lyase sequence is expressed in response to wounding and methyl jasmonate RNA is isolated from wounded leaves and leaves treated with methyl jasmonate (MJ). HPO lyase expression is also examined for induction by fungal attack.

Arabidopsis thaliana ecotype *No-O* is grown in soil under 16 hour light at 22° C at 65% relative humidity for three weeks. For wounding, each leaf is wounded once with a hemostat on the upper third of the leaf in line with the midvein. In each rosette, half of the leaves are wounded and half are not. These are referred to as local and systemic leaves respectively.

For treatment with MJ, plants are enclosed in airtight 9.25 L jars. Neat MJ (10 or 50 μl, Aldrich Co. Milwaukee, WI) is applied onto four cotton swabs and placed in the jars without directly touching the plants. Fresh MJ treated cotton swabs are replaced in the jar each time the jar is opened. For pathogen induction, spores of the fungus *Botrytis cinerea* are sprayed on the rosette leaves of 3 week old plants at a concentration of 10^6 spores per ml in 1% glucose.

Total RNA is isolated from the tissues using TRIzol reagent (Life Technologies) as described by the manufacturer and transferred to membrane as described above.

Expression of the HPO lyase mRNA is observed at 6 and 24 hours after wounding. High levels of HPO lyase induction are seen in leaves wounded with a hemostat. The induction is distinct after 6 hours of the treatment and the amount of HPO lyase mRNA increased at least until 24 hours. In the systemic leaves, induction of the HPO lyase mRNA is also evident. After 6 hours of treatment, the level of HPO lyase mRNA is almost the same as that of the local leaves, however, the amount increases only slightly afterwards and at 24 hours after treatment, the level is observably lower than that of the local leaves.

Plants treated with methyl jasmonate demonstrated a low expression of HPO lyase are observed in both 10µl and 50µl treatments at all time points examined.

Infection with the fungal pathogen *Botrytis cinerea* did not induce HPO lyase expression. However, necrotic lesions are observed on the leaf surfaces of treated plants 5 to 6 days post inoculation.

Example 4 Preparation HPO Lyase Expression Constructs

A set of constructs are prepared for transformation into either plant or bacterial hosts to further characterize the *Arabidopsis* HPO lyase-like sequence. The 5' RACE product in pCGN8091 was PCR amplified using the primers Alex2 (5'-CGGGATCCATGTTGTTGAGAACGATGGCGGCG-3') and Alex4 (5'-CAATCTCCGGCGTTCTCGTCG-3'). The Alex2 primer contains the restriction endonuclease site *Bam*HI for the convenient cloning of the PCR product into the pQE30 expression vector (Qiagen, Hilden, Germany) in frame with the ATG start codon of the vector. In addition to the oligonucleotide primers (0.2µM each), the PCR reaction mix contained 0.2 mM each of dATP, dCTP, dGTP and dTTP, 1.0% glycerol, 0.2 mM Tris-HCl (pH 8.3), 4.6 mM KCl, 1.5 mM EDTA, 15 µM dithiothreitol, 7.3 µgm/ml BSA, 1.1 mM KOAc and 0.1 units *Pfu* DNA polymerase (BRL Life Technologies, Gaithersburg, MD). The mixtures were amplified using the following conditions: 1 cycle of 95°C for 10 minutes; 30 cycles of 94°C for 20 seconds, 60°C for 30 seconds, and 72°C for 1.5 minutes; and, 1 cycle of 72°C for 7 minutes in a Perkin-Elmer 9800 thermocycler. The resulting PCR product was digested with *Bam*HI and *Hind*III and ligated into the vector pQE30 to create the vector

pCGN8099. The 3' terminus of the *Arabidopsis* HPO lyase was cloned into the *Hind*III site of pCGN8099 from pCGN8094 to create the *E. coli* expression vector pCGN8100

A binary vector for plant transformation, pCGN5138, was constructed from pCGN1558 (McBride and Summerfelt, (1990) Plant Molecular Biology, 14:269-276). The polylinker of pCGN1558 was replaced as an *Hind*III/*Eco*RI fragment with a polylinker containing unique restriction endonuclease sites, *Hind*III, *Sse*I/*Pst*I, *Not*I, *Bam*HI, *Swa*I, *Xba*I, *Pac*I, *Asc*I, and *Asp*718.

An antisense construct of the *Arabidopsis* HPO lyase-like nucleotide sequence was prepared for transformation of *Arabidopsis*. The nucleic acid sequence encoding the 5' 1000 bp nucleotides from pCGN8091 were cloned as an *Eco*RI fragment into the plasmid pBluescript II SK (Stratagene, La Jolla, CA) to create the vector pCGN8093. The 3' RACE product from pCGN8090 was cloned as a *Hind*III fragment into pCGN8093 to create a full length HPO lyase coding sequence in the plasmid pCGN8094. The *Kpn*I site of pCGN8094 was removed by digesting with *Kpn*I and filling in the site with Klenow fragment, and the HPO lyase coding sequence was cloned from this plasmid as a *Sma*I fragment into the *Stu*I site of pCGN8059. This yields the plasmid pCGN8101. The plasmid pCGN8059 contains a multiple cloning site downstream of the 35S promoter and the hsp70 leader sequence to allow for the cloning of sequences for expression from the 35S promoter sequence. This vector also contains the nopaline synthase transcription termination (nos 3') sequences (Fraley et al., *Proc. Natl. Acad. Sci* (1983) 80:4803-4807 and Depicker et al., *J. Molec. Appl. Genet.* (1982) 1: 562-573). The fragment containing the 35S promoter/hsp70 leader, antisense *Arabidopsis* HPO lyase sequence, and *nos*3' termination sequence was cloned from pCGN8101 as a *Not*I fragment into the same site of pCGN5138 to create the antisense expression construct pCGN8102.

Example 5 E coli Expression

The expression vector pCGN8100 was transformed into *E coli* (strain M15, Qiagen, Hilden, Germany) using a calcium chloride procedure described in Maniatis, *et al.* ((1989) *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). Transformed colonies were screened by Western immunoblot analysis for expression of the HPO Lyase protein using antibodies raised to the bell pepper HPO lyase as described in Shibata, *et al.* (1995) *Plant Cell Physiol.* 97:1059-1072.

Hydroperoxide lyase activity was determined by gas chromotography (GC) methods described by Matsui, *et al.* (1991), *Phytochemistry*, 30:2109-2113.

TABLE 1

Sample	Area	nmole	nmole/10min/mg
8100	24677	130	153
Control	4089	28	24

The results of the GC analysis shown in Table 1, demonstrates that the *Arabidopsis* HPO lyase-like sequence encodes a HPO lyase enzyme.

Example 6 Transformation of *Arabidopsis* with Antisense HPO Lyase Constructs

A variety of methods have been developed to insert a DNA sequence of interest into the genome of a plant host to obtain the transcription or transcription and translation of the sequence to effect phenotypic changes.

The plant binary constructs pCGN8101 are used in plant transformation to direct the expression of the antisense nucleic acid sequence of the *Arabidopsis* HPO lyase-like sequence from plant tissues.

Transgenic *Arabidopsis thaliana* plants may be obtained by *Agrobacterium*-mediated transformation as described by Valverkens *et al.*, (*Proc. Nat. Acad. Sci.* (1988) *85*:5536-5540), or as described by Bent et al. ((1994), *Science* 265:1856-1860), or Bechtold et al. ((1993), *C.R.Acad.Sci, Life Sciences* 316:1194-1199).

Example 7 Analysis of Transgenic Plants

Transgenic *Arabidopsis* plants containing pCGN8101 are analyzed for the decreased production of Hexenal by High Pressure Liquied Chromotography (HPLC) analysis of protein extracts as described in Shibata *et al.* (1995) *Plant Cell Physiol.* 36:147-156.

Transgenic plants overexpressing the *Arabidopsis* HPO lyase of the present invention may be screened using a photometric assay or by the HPLC assay which are both described in Shibata, *et al.* (1995) *supra*.

Example 8 Identification of Additional HPO lyase Sequences

Additional HPO lyase-like sequences are obtained from tomato and cucumber tissues. Total RNA was isolated from cucumber hypocotyls and tomato immature fruit tissue using TRIzol reagent (Gibco-BRL Life Technologies, Gaithersburg, MD) following the manufacturers protocol.

Complementary DNAs (cDNA) are obtained using the Marathon cDNA Amplification Kit (Clontech, Palo Alto, CA) following the manufacturers directions.

The sequences of HPO lyase from bell pepper (Matsui, *et al.* (1996) *supra*), banana (European Patent Application, Publication Number EP 0 801 133 A2) and *Arabidopsis* were aligned using ClustalW (http://www.clustalw.genome.ad.jp/) and seven conserved peptide sequences were identified (see Figure 7 for positions, Table 2 for a listing).

Table 2

	Peptide Sequence	Primer Name	Oligonucleotide Sequence
1	PGSYG	HPOL1S	5'-ATNCCNGGNWSNTAYGG-3'
2	QPLEEI	HPOL2S	5'-CARCCNYTNGARGARAT-3'
		HPOL2AS	5'-ATYTCYTCNARNGGYTG-3'
3	GFNAYGG	HPOL3S	5'- GGNTTYAAYGCNTWYGGNGG-3'
		HPOL3AS	5'-CCNCCRSANGCRTTRAANCC-3'
4	YQPLVM	HPOL4S	5'-TAYCARCCNYTNGTNATG-3'
		HPOL4AS	5'-CATNACNARNGGYTGRTA-3'
5	VFDEPE	HPOL5S	5'-GTNTTYGAYGANCCNGA-3'
		HPOL5AS	5'-TCNGGNTCRTCRAANAC-3'
6	NGPQTG	HPOL6AS	5'-CCNGTYTWNGGNCCRTT-3'
7	NKQCAAKD	HPOL7AS	5'-CYTTNGCNGCRCAYTGYTTRTT-3'

A set of synthetic oligonucleotides (Table 2) are synthesized for use in polymerase chain reactions with the cDNAs obtained above to identify sequences which are homologous to HPO lyase sequences. The PCR reactions are carried out using Advantage cDNA Polymerase Mix (Clonetech, Palo Alto,CA) using the reaction conditions according to the manufacturers protocol. The letter "S" in the oligonucleotide name designates a PCR primer designed to amplify the sense strand, or forward reaction primer. The letters "AS" designates a PCR primer designed to amplify the antisense strand, or reverse reaction primer. In the oligonucleotide sequence, the letters "N" represents an A, C, G, or a T, the letter "S"

represents a C or a G in that position, the letter "Y" represents a C or a T, and the letter "R" represents an A or a G in that position.

A single PCR product, of approximately 475 bp, was amplifed in reactions containing the primers 4HPOL3S and 11HPOL7AS, with the cDNAs obtained from both cucumber and tomato (described above). The 475 bp PCR product from tomato and cucumber were cloned into the plasmid pCR2.1TOPO (invitrogen) to yield the plasmids T15 and C15 respectively. In PCR reactions with 6HPOL4S and 11HPOL7AS, a single product, of approximately 200 bp, was obtained from amplification reactions with cDNA obtained from cucumber hypocotyl tissue. The 200 bp product was cloned into pCR2.1 TOPO (Invitrogen), to create the plasmid C17.

The nucleotide sequence of each PCR product was determined by automated sequencing. The sequences obtained are compared to nucleic acid and amino acid sequences of HPO lyase sequences from bell pepper, *Arabidopsis*, and banana leaf, as well as to DNA and amino acid sequences coding for allene oxide synthases from guayule ((1995) *J. Biol. Chem.* 270(15):8487-8494), flaxseed ((1993) *Proc Natl Acad Sci USA* 90(18):8519-8523) and *Arabidopsis*.

The results demonstrate that the T15 nucleic acid sequence is approximately 85% similar to the bell pepper HPO lyase DNA sequence and about 88% similar in the amino acid sequence. Furthermore, the T15 sequence is also at least about 55% similar to other HPO lyase nucleic acid sequences and at least about 57% similar in the amino acid sequence. In addition, the T15 amino acid sequence is only about 41% similar to the allene oxide synthase sequences. The C17 sequence also follows a similar pattern of similarity to the HPO lyase sequences. Thus, the T15 and C17 sequences encode proteins highly similar to HPO lyase.

However, the results of the sequence comparisons (Figure 8) demonstrate that the C15 nucleic acid sequence is between 50% and 54% similar to the other HPO lyase nucleic acid sequences and about 58% similar to the allene oxide synthase DNA sequences. Furthermore, the deduced amino acid sequence of C15 is between about 38% and 42% similar to the HPO lyase amino acid sequences and about 51% similar to the AOS amino acid sequences. Thus, the C15 sequence encodes a protein which is divergent from both the known HPO lyase sequences, and is more similar to allene oxide synthase sequences.

The nucleotide sequence of each PCR product was determined by automated sequencing. The sequences obtained are used to search Genbank. Search results identify the

sequences from T15 and C17 as being similar to HPO lyase sequences, while the sequence of C15 is similar to allene oxide synthase sequences.

In order to obtain full length coding sequence for T15, C15 and C17, RACE PCR reactions are employed using the Marathon cDNA Amplification kit (Clontech) according to the manufacturers protocol, and the oligonucleotides shown in Table 3.

Table 3

1KMC10-1: 5'-CGGTGGAGATCCTCGCCACCGGTGCCGACCC-3'
2KMC10-2: 5'-CTTCCTTCACGGTTGTCCTCACTTCCTCCGCCAG-3'
3KMC17-1: 5'-TCCAGCAGCGCTGCCCCTTTCTCTCCCCGG-3'
4KMC17-2: 5'-CACTGTTTGTTCTTCTCGCTCGGTGTCCCCG-3'
5KMC10-3: 5'-GGGTCGGCACCGGTGGCGAGGATCTCCACCG-3'
6KMC10-4: 5'-CTGGCGGAGGAAGTGAGGACAACCGTGAAGGAAG-3'
7KMC17-3: 5'-CCGGGGAGAGAAAGGGGCAGCGCTGCTGG-3'
8KMC17-4: 5'-CGGGGACACCGAGCGAGAAGAACAAACAGTG-3'
9KMT15-1: 5'-GACTTGGTACTGGTGGACTAAGCCTAAGTGTTTC-3'
10KMT15-2: 5'-GGCTGATAACCACAAAGAAGCTCCCCTTTC-3'
11KMT15-3: 5'-GAAACACTTAGGCTTAGTCCACCAGTACCAAGTC-3'
12KMT15-4: 5'-GAAAGGGGAGCTTCTTTGTGGTTATCAGCC-3'

PCR products from the amplification reactions with DNA obtained from tomato and cucumber are cloned into pCR2.1 TOPO. The sequences of the 5' and 3'-RACE products from tomato (pCGN8303 (5' RACE) and pCGN8304 (3' RACE)), cucumber, C15 (pCGN8302 (5'RACE) and pCGN8306 (3' RACE)) and C17 (pCGN8301 (5' RACE) and pCGN8307 (3' RACE)) are sequenced and aligned with the respective sequences obtained from pCGN8305, pCGN8309, and pCGN8308 to obtain preliminary full length sequences corresponding to a tomato HPO lyase-like sequence (Figure 4), a cucumber HPO lyase-like sequence (Figure 6) and a cucumber allene oxide synthase-like sequence (Figure 5).

Example 9 Preparation of Expression Constructs

A set of constructs are prepared for transformation into either plant or bacterial hosts to further characterize the novel sequence from cucumber. To create a full length coding sequence for the cucumber (C15) allene oxide synthase-like sequence, the sequences from the

5' RACE (pCGN8302) and 3' RACE (pCGN8306) were PCR amplified and combined at a unique restriction endonuclease site.

The 5' C15 sequence is amplified using primers (4KMC15ES1 5'-CGGGATCCATGGCTTCTTCCTCCCCTGAACTTC-3' and 5KMC15EAS2 5'-TGCCGACCCATTTCAGTATAGTGGG-3') in PCR amplification reactions described above. The primer 4KMC15EAS1 amplifies from the 5' region and contains the start codon (ATG), and a *Bam*HI site. The 3' C15 sequence is amplified using the AP1 primer provided in the Marathon Kit (BRL-Lifetechnologies, Gaithersburg, MD) and the primer 6KMC15ES3 (5'- TTCACACCATTCCCCTGCCTTTCTTCCC-3'). The sequence of the C15 full length clone is shown in Figure 6.

A. Bacterial Expression Construct

The 5' RACE PCR amplification product is digested with *Bam*HI and *Xba*I (unique site endogenous to the C15 sequence) and cloned into the expression vector pQE30 (Invitrogen) with the amplification product of the 3' RACE PCR reaction digested with *Xba*I and *Sma*I. This construct provides a full length encoding sequence of the C15 cDNA in the *E. coli* expression vector to create the vector pCGN8333. The full length sequence is also cloned into the plasmid pUC119 to create the vector pCGN8334.

B. Plant Expression Construct

A binary vector for plant transformation, pCGN5138, was constructed from pCGN1558 (McBride and Summerfelt, (1990) Plant Molecular Biology, 14:269-276). The polylinker of pCGN1558 was replaced as an *Hind*III/*Eco*RI fragment with a polylinker containing unique restriction endonuclease sites, *Hind*III, *Sse*I/*Pst*I, *Not*I, *Bam*HI, *Swa*I, *Xba*I, *Pac*I, *Asc*I, and *Asp*718.

The full length coding sequence of C15 is cloned to be expressed from the plant constitutive promoter 35S for expression in plants. The expression cassette is cloned into the binary vector pCGN5138 to create the vector pCGN8337.

Example 10 Expression of cucumber C15 in *E. coli*

The expression vector pCGN8333 was transformed into *E coli* (strain M15, Qiagen, Hilden, Germany) using a calcium chloride procedure described in Maniatis, *et al.* ((1989) *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). Transformed colonies were screened by Western immunoblot analysis for expression of the HPO Lyase protein using antibodies raised to the

bell pepper HPO lyase as described in Shibata, *et al.* (1995) *Plant Cell Physiol.* 97:1059-1072.

Hydroperoxide lyase activity was determined by spectrophotometric and gas chromotography (GC) methods described by Matsui, *et al.* (1991), *Phytochemistry*, 30:2109-2113, using both linolenic acid 13-hydroperoxide and linolenic acid 9-hydroperoxide as substrates.

The results of the gas chromotography assay (Figure 9) demonstrate that the protein encoded by the cucumber C15 sequence has greater activity toward linolenic acid 9-hydroperoxide (Figure 9B) substrates than linolenic acid 13-hydroperoxide substrates (Figure 9A). The results of the spectrophotometric assays further demonstrate the preference of the protein encoded by cucumber HPO lyase nucleic acid sequence for 9-Hydroperoxide substrates. The results of the spectrophotometric assay are presented in Figure 10.

Thus, the cucumber C15 sequence represents the first known cloning of a nucleic acid sequence encoding a 9-hydroperoxide lyase.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claim.

Claims

What is Claimed is:

1. An isolated nucleotide sequence of the *Arabidopsis* HPO lyase.

2. The nucleotide sequence of Claim 1 comprising the sequence shown in Figure 1.

3. A construct comprising the nucleotide sequence of Claim 1 joined to a heterologous nucelotide sequence.

4. An isolated nucleotide sequence of the tomato HPO lyase.

5. The isolated nucleotide sequence of Claim 4comprising the sequence shown in Figure 4.

6. A construct comprising the nucleotide sequence of Claim 4 oined to a heterologous nucelotide sequence.

7. An isolated nucleotide sequence of the cucumber HPO lyase.

8. The isolated nucleotide sequence of Claim 7 comprising the sequence shown in Figure 5.

9. A construct comprising the nucleotide sequence of Claim 7 oined to a heterologous nucelotide sequence.

10. A method for obtaining an isolated nucleotide sequence comprising an HPO lysase, said method comprising obtaining amplification products from a PCR reaction using an oligonucleotide selected from the group consisting of,

5'-ATNCCNGGNWSNTAYGG-3';

5'-CARCCNYTNGARGARAT-3';

5'-ATYTCYTCNARNGGYTG-3';

5'- GGNTTYAAYGCNTWYGGNGG-3';

5'-CCNCCRSANGCRTTRAANCC-3';

5'-TAYCARCCNYTNGTNATG-3';

5'-CATNACNARNGGYTGRTA-3';

5'-GTNTTYGAYGANCCNGA-3';

5'-TCNGGNTCRTCRAANAC-3';

5'-CCNGTYTWNGGNCCRTT-3'; and

5'-CYTTNGCNGCRCAYTGYTTRTT-3'.

11. A PCR amplification product comprising an oligonucleotide selected from the group consisting of:

1KMC10-1: 5'-CGGTGGAGATCCTCGCCACCGGTGCCGACCC-3';

2KMC10-2: 5'-CTTCCTTCACGGTTGTCCTCACTTCCTCCGCCAG-3';

3KMC17-1: 5'-TCCAGCAGCGCTGCCCCTTTCTCTCCCCGG-3';

4KMC17-2: 5'-CACTGTTTGTTCTTCTCGCTCGGTGTCCCCG-3';

5KMC10-3: 5'-GGGTCGGCACCGGTGGCGAGGATCTCCACCG-3';

6KMC10-4: 5'-CTGGCGGAGGAAGTGAGGACAACCGTGAAGGAAG-3';

7KMC17-3: 5'-CCGGGGAGAGAAAGGGGCAGCGCTGCTGG-3';

8KMC17-4: 5'-CGGGGACACCGAGCGAGAAGAACAAACAGTG-3';

9KMT15-1: 5'-GACTTGGTACTGGTGGACTAAGCCTAAGTGTTTC-3';

10KMT15-2: 5'-GGCTGATAACCACAAAGAAGCTCCCCTTTC-3';

11KMT15-3: 5'-GAAACACTTAGGCTTAGTCCACCAGTACCAAGTC-3'; and

12KMT15-4: 5'-GAAAGGGGAGCTTCTTTGTGGTTATCAGCC-3'.

12. A construct comprising a DNA sequence capable of directing transcription in a in a host cell linked to an nucleotide sequence encoding an HPO lyase,

wherein said HPO lyase comprises an amino acid sequence selected from the group consisting of,

HPOL1S (PGSYG);

HPOL2S (QPLEEI);

HPOL2AS;

HPOL3S (GFNAYGG);

HPOL3AS;

HPOL4S (YQPLVM);

HPOL4AS;

HPOL5S (VFDEPE);

HPOL5AS;

HPOL6AS (NGPQTG); and

HPOL7AS (NKQCAAKD);

and wherein said HPO lysase is not a banana or bell pepper HPO lyase.

13. A method for increasing the resistance of a plant to a plant pathogen comprising expressing an HPO lyase from a construct according to any one of Claims 3, 6, 9 or 12, wherein said HPO lysase encoding sequence is linked to DNA sequence capable of directing expression in a plant cell.

14. A method for increasing the volatile composition of a plant comprising expressing an HPO lyase from a construct according to any one of Claims 3, 6, 9 or 12, wherein said HPO lysase encoding sequence is linked to DNA sequence capable of directing expression in a plant cell.

15. An isolated nucleic acid sequence coding a hydroperoxide lyase with activity toward fatty acid 9-hydroperoxides.

16. The sequence according to Claim 15 wherein said sequence has activity toward linolenic acid 9-hydroperoxide.

17. The sequence according to Claim 15 wherein said sequence is obtained from a plant source.

18. The sequence according to Claim 15 wherein said sequence is obtained from cucumber.

19. The sequence according to Claim 15 wherein said sequence is obtained from cucumber hypocotyl.

20. The sequence according to Claim 15 comprising the sequence shown in Figure 6.

21. A construct comprising a promoter functional in a host cell, a sequence encoding a 9-hydroperoxide lyase, and a transcriptional termination sequence.

22. A construct according to Claim 21herein said 9-hydroperoxide lyase sequence is isolated from a plant.

23. A construct according to Claim 21herein said 9-hydroperoxide lyase sequence is isolated from cucumber.

24. A method for increasing the resistance of a plant to a plant pathogen comprising expressing an HPO lyase from a construct according to Claim 21wherein said HPO lysase encoding sequence is linked to DNA sequence capable of directing expression in a plant cell.

25. A method for increasing the volatile composition of a host cell comprising expressing an HPO lyase from a construct according to Claim 21wherein said HPO lysase encoding sequence is linked to DNA sequence capable of directing expression in a host cell.

26. The method according to Claim 25 wherein said host cell is a plant cell.

27. The method according to Claim 25 wherein said host cell is a microbial cell.

28. The method according to Claim 27 wherein said microbial cell is a yeast cell.

29. The method according to Claim 25 further comprising the step of harvesting material from said host cell and concentrating said volatile composition in said harvested material.

ctagaattca gcggccgctg aattccgaag agataatttc caaaaacaag aaaaacctct
aagctcaaaa gatgttgttg agaacgatgg cggcgacttc cccgcggcca ccaccgtcaa
catccctaac atctcagcag ccaccatcac ccccctcaca gcttcccctc cgtacaatgc
cgggatcgta cggctggccg ttggttggac cattatcgga ccgtttagat tacttctggt
tccaaggacc cgataagttt ttccggacaa gagctgagaa gtataagagc actgtgttcc
gtacaaatat tcctccgacg tttcctttct tcggcaacgt taaccctaac atcgtcgccg
ttcttgacgt caagtctttt agccatcttt ttgacatgga tctagttgat aaaagagatg
ttctcatcgg agacttccgg cctagccttg ggttctacgg cggcgttcgt gttggagttt
atctcgacac tactgagcca aagcacgcca agataaaagg tttcgctatg gaaacactaa
aacgaagctc aaaagtatgg ctacaagagc ttcgttcaaa cctaaacatt ttctggggaa
caatcgaatc cgaaatctcc aaaaacggtg ccgcttcata tatcttccct ctccaacgtt
gcatcttcag tttcctctgc gcctctctcg ccggcgttga cgcttccgta tcgccggaca
tcgctgagaa cggttggaaa acaatcaata cttggcttgc gttgcaagtt attcccactg
ctaaacttgg cgtagttcct cagcctcttg aagagatttt acttcatact tggccttatc
cttctctctt aatcgccgga aattacaaaa agctttacaa tttcatcgac gagaacgccg
gagattgtct ccggttaggt caagaagaat tcgggttgac ccgagatgag gctattcaaa
atcttctctt tgttttaggt tttaatgcct acgggggctt ttccgtcttc ttaccttctt
tgatcgggag aataaccggc gacaattccg gtttacagga gaggattaga accgaagtca
ggagagtttg cggatccggg tcggatctta atttcaagac ggttaacgaa atggagctgg
ttaaatccgt ggtttacgaa acgctgcgtt ttagtcctcc ggttccgctg caattcgcac
gtgcgaggaa agattttcag ataagttcac acgatgctgt ttttgaggtc aagaaaggtg
agcttctttg tggttatcag ccgcttgtga tgagagacgc taatgttttt gacgaaccgg
aggaatttaa accggaccgg tatgtaggtg agaccgggtc tgaattgctg aattatctct
actggtctaa cggtccacaa accggtaccc cgagcgcgtc gaacaaacag tgtgcagcta
aggacattgt cactctcacg gcttccttgc tcgttgccga tttatttctc cggtatgata
cgattactgg tgactccggt tcaattaaag ctgttgttaa agctaaataa gttaaagatt
gattagtgtt taatgttatg taatcaatct tcaacggttt catttgttcg gagtgtataa
attatatgta aactaaaata atccgaattg aaactcaaat tgttatgcaa aaaaaaaaaa
aaaaaaa

Figure 1

[57.0% / 470 aa]

```
AraHPOL MLLRTMAATS PRPPPSTSLT SQQPPSPPSQ LPLRTMPGSY GWPLVGPLSD
                                        *   ** *  **** * ** *** *
CaHPOL                MIPIMSS APLSTATPIS LPVRKIPGSY GFPLLGPLWD

AraHPOL RLDYFWFQGP DKFFRTRAEK YKSTVFRTNI PPTFPFFGNV NPNIVAVLDV
        **** ***     **  * ** * *******  ** ****  * *** ******
CaHPOL  RLDYNWFQKL PDFFSKRVEK YNSTVFRTNV PPCFPFFLGV NPNVVAVLDV

AraHPOL KSFSHLFDMD LVDKRDVLIG DFRPSLGFYG GVRVGVYLDT TEPKHAKIKG
        *** *****   * *  ** * ** **    *   **  ****  ****  **
CaHPOL  KSFAHLFDME IVEKANVLVG DFMPSVVYTG DMRVCAYLDT SEPKHTQIKN

AraHPOL FAMETLKRSS KVWLQELRSN LNIFWGTIES EISKNGAASY IFPLQRCIFS
        *    ***** * *   *    *    ** **   **    *     **   *
CaHPOL  FSLDILKRSS KTWVPTLVKE LDTLFGTFES DLSKSKSASL LPALQKFLFN

AraHPOL FLCASLAGVD ASVSPDIAEN GWKTINTWLA LQVIPTAKLG VVPQPLEEIL
        *      * *  * ** **   *      ***  *  **   * *  *******
CaHPOL  FFSLTFLGAD PSASPEIANS GFAYLDAWLA IQLAPTVSIG VL-QPLEEIF

AraHPOL LHTWPYPSLL IAGNYKKLYN FIDENAGDCL RLGQEEFGLT RDEAIQNLLF
         *   **  *   * * **   *    *   *    *  * **   *** ****
CaHPOL  VHSFSYPYFL VRGGYEKLIK FVKSEAKEVL TRAQTDFQLT EQEAIHNLLF

AraHPOL VLGFNAYGGF SVFLPSLIGR ITGD-NSGLQ ERIRTEVR-R VCGSGSDLNF
         ***** ***   *** * *       *   * *  * ***   *       * *
CaHPOL  ILGFNAFGGF TIFLPTLLGN LGDEKNAEMQ EKLRKEVREK VGTNQENLSF

AraHPOL KTVNEMELVK SVVYETLRFS PPVPLQFARA RKDFQISSHD AVFEVKKGEL
          * *****  * *** ** * **** * *** ****  ****  * * *****
CaHPOL  ESVKEMELVQ SFVYESLRLS PPVPSQYARA RKDFMLSSHD SVYEIKKGEL

AraHPOL LCGYQPLVMR DANVFDEPEE FKPDRYVGET GSELLNYLYW SNGPQTGTPS
        *********  *  ******  *   *   * * ****** * ******* *
CaHPOL  LCGYQPLVMK DPKVFDEPEK FMLERFTKEK GKELLNYLFW SNGPQTGSPT

AraHPOL ASNKQCAAKD IVTLTASLLV ADLFLRYDTI TGDSGSIKAV VKAK
         *********  ******* * *  *  **     ***   *  **
CaHPOL  ESNKQCAAKD AVTLTASLIV AYIFQKYDSV SFSSGSLTSV KKAC
```

Figure 2

[39.2% / 485 aa]

```
AraHPOL                              MLLRTMAAT SPRPPPSTSL TSQQPPSPPS
                                       *                 *
AraAOS  MASISTPFPI SLHPKTVRSK PLKFRLVTRP IKASGSETPD LTVATRTGSK

AraHPOL QLPLRTMPGS YGWPLVGPLS DRLDYFWFQG PDKFFRTRAE KYKSTVFRTN
         ** *  **  ** * ***   ** ***  **     **  *   ** *** * *
AraAOS  DLPIRNIPGN YGLPIVGPIK DRWDYFYDQG AEEFFKSRIR KYNSTVYRVN

AraHPOL IPPTFPFFGN VNPNIVAVLD VKSFSHLFDM DLVDKRDVLI GDFRPSLGFY
         **   *     **  ** **  ***  ***  * * * *    *   **
AraAOS  MPPG-AFIAE -NPQVVALLD GKSFPVLFDV DKVEKKDLFT GTYMPSTELT

AraHPOL GGVRVGVYLD TTEPKHAKIK GFAMETLKRS SKVWLQELRS NLNIFWGTIE
         ** *  ***   **** * *        ** *         *          *
AraAOS  GGYRILSYLD PSEPKHEKLK NLLFFLLKSS RNRIFPEFQA TYSELFDSLE

AraHPOL SEISKNGAAS YIFPLQRCIF SFLCASLAGV DASVSPDIAE NGWKTINTWL
         *                 *     **  *     * *              *  *
AraAOS  KEAFPLRESG FRRFQRRNRL LFLGSSFL-R DESRRYKLKA DAPGLITKWV

AraHPOL ALQVIPTAKL GVVPQPLEEI LLHTWPYPSL LIAGNYKKLY NFIDENAGDC
           *      * *  **  * **   *   *    * **  *     *
AraAOS  LFNLHPLLST G-LPRVIEEP LIHTFSLPPA LVKSDYQRLY EFLRIR-GEI

AraHPOL LRLGQEEFGL TRDEAIQNLL FVLGFNAYGG FSVFLPSLIG RITGDNSGLQ
         *       *  * **  *** *   **  **       *     **
AraAOS  L-VEADKLGI SREEATHNLL FATSFNTWGG MKILFPNMVK RIGPGGHQVH

AraHPOL ERIRTEVRRV C-GSGSDLNF KTVNEMELVK SVVYETLRFS PPVPLQFARA
         *   * * *      *  *          *** * ***** ***  ***  *  **
AraAOS  NRLAEEIRSV IKSNGGELTM GAIEKMELTK SVVYECLRFE PPVTAQYGRA

AraHPOL RKDFQISSHD AVFEVKKGEL LCGYQPLVMR DANVFDEPEE FKPDRYVGET
         **    *** * * ** **  * *****  * *   **   * * * * ***
AraAOS  KKDLVIESHD AAFKVKAGEM LYGYQPLATR DPKIFDRADE FVPERFVGEE

AraHPOL GSELLNYLYW SNGPQTGTPS ASNKQCAAKD IVTLTASLLV ADLFLRYDTI
         *  **    * **** * **    ***** **  * * * * *    * ***
AraAOS  GEKLLRHVLW SNGPETETPT VGNKQCAGKD FVVLVARLFV IEIFRRYDSF

AraHPOL TGDSGSIKAV VKAK
           *
AraAOS  DIEVGTSPLG SSVNFSSLRK ASF
```

Figure 3

ttgataatga tattcagctt accacccaac gccctcaacc ttgactagcc ctccctcaaa
ctctcctttt tcaccaatct cactctccaa atagtatttt ctaggtttca tagtgataag
cacatccaca caagagctat gaaatttctg aaacaaacgt tacaatagac acacggcgta
tttatttccc tcaagcaatc gtatcccctt agcagatcga aaaaaaaaga agataccgat
aatgaattct gctcctctat caactccggc gccggttact ctgcccgtcc gttcgattcc
gggcagctac ggattgccgt tagtagggcc aatcgcggat cgattagact acttctggtt
ccaaaaaccg gaaaacttct tcaccaagag aatggaaaag cacaagagca cggttttcag
aacgaacgtg ccgccgtgtt ttccgttttt cggtagtgtg aatccaaatg tggtggcggt
tttagatgtg aaatcgtttt cgcatctgtt tgatatggag attgtagaga aagctaatgt
gcttgttggg gatttcatgc ccagtgttgt ttatactgga gatatgcgtg tttgtgctta
tctcgatact tctgaaccta aacatgccca gattaagaat ttttcacagg atattctaaa
aagaggctca aaaacatggg tgcctacact acttaaagag cttgatacaa tgtttacaac
ttttgaagca gatctttcaa aatccaatac agcttctctt cttcctgcac tccaaaaatt
cctctttaac ttcttctccc tcactatcct cggcgctgat ccatctgtct ccccagaaat
cgccaattct ggctacatct ttcttgattc atggctagct attcaattag cacctactgt
cagcattggt gtccttcaac ctcttgaaga aattttagtc cattcttttg catacccttt
ttttcttgtc aaaggtaatt atgaaaaact cgttcaattc gtcaaaaatg aagccaagga
agttttaagc agggcacaaa cagagtttca actcacagaa caagaagcca ttcataacct
tttgttcatt cttgggttca atgcttttgg tggcttctcc attttcttgc caactctttt
gggaaatctt ggagatgaga aaaacgcgga tatgcaagag aaactgagaa aagaagtgag
agacaaagtc ggcgtaaatc cagaaaactt gagttttgaa agtgttaaag aaatggaact
tgttcagtct tttgtttatg aaacacttag gcttagtcca ccagtaccaa gtcaatatgc
aagagcaaga aaagatttta aactgagttc acatgattca gtttacgaaa tcaagaaagg
ggagcttctt tgtggttatc agcctttagt tatgaaagat ccaaaggtgt ttgatgaacc
tgaaaagttt gtgttggaga gatttacaaa ggaaaaaggg aaagaattgc tgaattattt
gttttggtct aatggtccac aaactgggag acctactgaa tcaaacaagc aatgtgctgc
taaagatatg gttactctaa ctgcttcttt gattgtggct tacattttcc aaaagtacga
ttctgtgagt ttctcatctg gttcactcac atctgtgaaa aaggctagtt gagttctggg
ctgtgtagct gcaaaatcta agcttgctag acatgttgat cttaagtctc ggtggctggc
ttttccttcc tccccctatt gtgtaatata atttcacaaa taaaagttca acatgagtgt
acttattttc tgttctgact tatgactatt gtctatcctc ttacaaatta agttggaaca
tgaatctaag ttataattac aaaaaaaaaa aaaaaaaaaa a

Figure 4

ctcctcttct ctctaccgga aaagttcaac caacctctcc cctcccttc aaattcatac
tcccacatct ctcaaatatc gatgcacctc accatgacct ccggcggaat gccttccata
ccttcatcta tttccccccc gccggtctct ttacctctca gaaatatccc cggcagctac
ggtttgccgc tgctgggatc catcggtgac cgactcgatt actactggtt tcaaggaccc
gataagttct tcaggactcg tatggagaag aatcgaagca cggttttcag aaccaatgtt
cctccgtctt tccctttcat ctccgctgat cccagagtag ttgcggttct ggattgcaag
tcctttgcgc atctatttga catggaaatc gtggagaaaa ataatgttct ggtcggtgat
ttcatgccca gcatcagttt caccggaaac atgagagtgt gtgcttattt ggatacctcg
gaacccaatc actccaaggt aaagaacttc attacagaca ttctacggcg gagctcaagg
atatggatat cggagttgga atccaaccta tcgacgatgt gggacggcat tgaattagaa
atggcaaaga acaaacaatc gggtttcaga aactttctcc aaccagctct tttcaatttc
ttctccaaaa ctctcgccgg cgccgacact gcaaaatcac cggaagtggc aaaatccgga
tacatagacg tcataatctg gctcggcctc cagctcgttc ccaccatcca catcggcgtt
ctccagcccc ttgaagaaat attccttcac tctttccgct taccctttt ccccattgct
tcccgctacc aaagactcta cgatttcttc caaaaagaag gggcagaagt tgttgagcga
ggtgtcaccg agttcgggtt aacaaaggaa gaagccattc acaatctcat cttcaccatg
ggcttcaacg cctatggtgg cttcagtctc ttcttccctg ttctacttga ccgtatactc
aatgataaaa ccggtctaca acaaagaatc ctcaaggaag tcaggtcaaa atccagctct
ggtctgacct tcgagtcagt caaggagatg gacctcatct actccatcgt ctacgagaca
ctccggcttg acccccagt tccctcccaa tatgccagag ccagaagatt tcaagctaac
ttcctacaat tcaacgtata acatcaagaa aggggaactg ctttgtgggt atcagccgct
ggtgatgaga gatccagagg tgttcgatga accggaggcg tttaatccag accgattccg
gggagagaaa ggggcagcgc tgctggatta tttattttgg tcgaatgggc cgcaaacggg
gacaccgagc gagaagaaca aacagtgtgc tgggaaggat ttagtggtgc tgacgggggt
ggtgtttgtg gcttacatat ttaaaaggta tgattcgatt gctggggaag gaggttccat
tacagctttt caaagggcta attgaaaagt gaaagtgaat tatatataca ttatgtatcg
tttattgaga gatgcacttt tcttgttcaa tggcttcttt cttatgtatg agtttggagc
caaaatggga agatttggaa tgaataaatc aataaaatca agatttactt taccttgtaa
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aa

Figure 5

ccattctctt ccaacgtgaa gataagaaac cttttgttta ctttttgtac gatcacaggt
cacagcaatg gcttcttcct cccctgaact tcctctcaaa cccattcccg gtggctatgg
cttccccttc ctcggtccca tcaaagaccg ttacgattac ttctatttcc aaggtagaga
cgaattcttc cgttcccgaa ttaccaaata caactccacc gtcttccacg ccaacatgcc
accgggcccc ttcatctcct ccgattccag agtcgttgtc ctcctcgatg ccctcagttt
tcccatcctc ttcgacacca ccaaagtcga gaaacgcaac attctcgacg gaacttacat
gccctccttg tccttcaccg gcggcattcg cacctgtgct tatttggacc catcggaaac
agagcacact gttctcaaac gcctctttct ctcctttctc gcttctcacc atgacaggtt
catcccctctg tttcgaagct ccttgtctga gatgtttgtt aagcttgaag ataaactcgc
cgacaaaaat aagatcgctg atttcaactc gattagtgat gccgtgtcgt ttgattatgt
tttccgttta ttctccgatg gaacccctga ttcgacatta gctgctgatg gacctggaat
gttcgattta tggcttgggc ttcaacttgc cccattggct tccattggcc ttcccaaaat
tttctctgtt tttgaagatc tcattattca caccattccc ctgcctttct tcccagtcaa
gagtcgttac aggaagcttt ataaagcgtt ttactcctcc tctggctcat ttctagacga
agcagagaaa caggggatag acagagagaa agcatgtcac aatttagtgt ttcttgctgg
attcaacgca tacgggggaa tgaaagtcct ttttcccact atactgaaat gggtcggcac
cggtggcgag gatctccacc gtaaactggc ggaggaagtg aggacaaccg tgaaggaaga
agggggactg actttctccg ccttggagaa aatgagtctg ctgaagtcgg tcgtgtacga
agctctgagg atcgaaccgc cggtgccgtt ccagtacggg aaagcgaagg aggatatcgt
gattcagagc cacgattctt gtttcaagat caaaaaaggg gagacgattt ttggttatca
gccgtttgct actaaagatc cgaagatttt taaggactcg gagaagttcg tgggcgatag
gttcgtggga gaagaagggg agaagctttt gaagtatgtt tactggtcaa atgagcggga
gacggtggag ccgacggcgg agaacaagca gtgtccgggg aagaatctgg tggtgatgat
gggcaggatt attgtggtgg aattcttcct ccgttatgat acgttcactg tggacgtcgc
agatttggcg ctgggcccgg cggtgaagtt caagtccttg accagagcca ccgcttcggt
ttagaagcta atgacaaaat tagtttttaa tcattttaca gaataaagga agggtgttcg
tgattggttg atgcagggaa gctgtggtgg acttgagcac acatgacaca tcattgattt
gttggggtta taaagtaatt tcctcgtgat ccgcgtcgtt tttaatttga gatctcattg
tgtgttgtaa cccgcgggtg atcttatttt tatagtttct tttttctcaa ctatgctcca
atttttaaat taaataaata ccaaatatca acttcataga caaaaaaaaa aaaaaaaaaa
aaaaaaaaaa aaa

Figure 6

```
Ca   ---------------MIPIMSSAPLSTATPISLPVRKIPGSYGFPLLGPLWDRLDYNWFQKLPDFFSKRVEKYNSTVFRTNVPPCFPFFLGVNPNVVAVLDVKSFAHLFDMEIVEKANVLVG
At   MLLRTMAATSPRPPPSTSLTSQQPPSPPSQL--PLRTMPGSYGWPLVGPLSDRLDYFWFQGPDKFFRTRAEKYKSTVFRTNIPPTFPFFGNVNPNIVAVLDVKSFSHLFDMDLVDKRDVLIG
Ms   ---------------MAMMWSSASATAVT--TLPTRPIPGSYGPPLVGPLKDRLDYFWFQGPETFFRSRMATHKSTVFRTNMPPTFPFFVGVDPRVVTVLDCTSFSALFDLQVVEKKNILIG
                                        PRIMER 1

Ca   DFMPSVVYTGDMRVCAYLDTSEPKHTQIKNFSLDILKRSSKTWVPTLVKELDTLFGTFESDLS-KSKSASLLPALQKFLFNFFSLTFLGADPSASPEIANSGFAYLDAWLAIQLAPTVSI
At   DFRPSLGFYGGVRVGVYLDTTEPKHAKIKGFAMETLKRSSKVWLQELRSNLNIFWGTIESEIS-KNGAASYIFPLQRCIFSFLCASLAGVDASVSPDIAENGWKTINTWLALQVIPTAKL
Ms   DYMPSLSFTGDTRVVVYLDPSEPDHARVKSFCLQLLRRGAKTWVSSFLSNLDVMLATIEQGIA-KDGSAGLFGPLQKCIFAFLCKSIIGADPSVSPDVGENGFVMLDKWLALQLLPTVKV

Ca   G-VLQPLEEIFVHSFSYPYFLVRGGYEKLIKFVKSEAKEVLTRAQTDFQLTEQEAIHNLLFILGFNAFGGFTIFLPTLLGNLGDEKNAEMQEKLRKEVREKVGTNQEN-LSFESVKEM
At   GVVPQPLEEILLHTWPYPSLLIAGNYKKLYNFIDENAGDCLRLGQEEFGLTRDEAIQNLLFVLGFNAYGGFSVFLPSLIGRITGDN-SGLQERIRTEVRRVCGSG-SD-LNFKTVNEM
Ms   GAIPQPLEEILLHSFPLPFFLVSRDYRKLYEFVEKQGQEVVRRAETEHGLSKHDAINNILFVLGFNAFGGFSVFFPTLLTTIGRDK-TGLREKLKDEVRRVMKSRGEKRPSFETVREM
        PRIMER 2                                                                PRIMER 3

Ca   ELVQSFVYESLRLSPPVPSQYARARKDFMLSSHDSVYEIKKGELLCGYQPLVMKDPKVFDEPEKFMLERFTKEKGKELLNYLFWSNGPQTGSPTESNKQCAAKDAVTLTASLIVAYIFQK
At   ELVKSVVYETLRFSPPVPLQFARARKDFQISSHDAVFEVKKGELLCGYQPLVMRDANVFDEPEEFKPDRYVGETGSELLNYLYWSNGPQTGTPSASNKQCAAKDIVTLTASLLVADLFLR
Ms   ELVRSTVYEVLRLNPPVPLQUGRARTDFTLNSHDAAFKVEKGELLCGYQPLVMRDPAVFDDPETFAPERFMGS-GKELLKYVFWSNGPETGTPTPANKQCAAKDYVVETACLLMAEIFYR
                                                  PRIMER 4  PRIMER 5                    PRIMER 6    PRIMER 7

Ca   YDSVSFSSGSLTSVKKAC---
At   YDTITGDSGSIKAVVKAK---
Ms   YDEFVCADDAISVTKLDRAREWE
```

Figure 7

1. nucleotide

	Ca HPOL	Le HPOL	C17	At HPOL	Ms HPOL	C15	Li AOS	At AOS
Ca HPOL		84.6	59.9	59.7	55.0	50.0	52.4	52.6
Le HPOL	84.6		59.4	59.5	55.6	50.9	49.9	53.1
Cs17 HPOL	60.0	59.5		61.0	58.6	52.9	54.5	52.5
At HPOL	59.7	59.5	61.0		58.3	52.0	52.1	51.7
Ms HPOL	55.1	55.6	58.7	58.3		54.1	56.4	51.7
Cs15 HPOL	50.0	50.7	52.8	52.1	53.9		58.1	57.6
Li AOS	52.1	49.7	54.5	52.3	56.5	58.3		61.8
At AOS	52.5	53.2	52.5	51.7	51.8	57.8	62.3	

2. amino acid

	Ca HPOL	Le HPOL	C17	At HPOL	Ms HPOL	C15	Gu AOS	Li AOS	At AOS
Ca HPOL		88.2	59.5	57.0	54.1	39.6	41.0	40.3	40.0
Le HPOL	88.2		61.4	58.7	57.3	39.4	41.6	40.5	40.4
Cs17 HPOL	59.5	61.4		56.2	53.6	41.7	43.3	41.9	41.1
At HPOL	57.0	58.7	56.2		56.1	38.3	41.1	41.1	39.2
Ms HPOL	54.1	57.3	53.6	56.1		41.9	44.5	44.5	43.2
Cs15 HPOL	39.6	39.4	41.7	38.4	41.9		51.6	51.2	51.4
Gu AOS	41.0	41.6	43.3	41.1	44.5	51.6		65.9	60.3
Li AOS	40.3	40.5	41.9	41.1	44.5	51.2	65.9		58.2
At AOS	40.0	40.4	41.1	39.2	43.2	51.4	60.3	58.2	

Ca HPOL; bell pepper fruit HPOL, Le HPOL; tomato fruit HPOL, CsC17 HPOL; cucumber hypocotyl HPOL (pseudo gene)
At HPOL; Arabidopsis inflorecence HPOL, Ms HPOL; banana leaf HPOL, Cs15 HPOL; cucumber hypocotyl 9-HPOL
Gu AOS; guayule AOS, Li AOS; flaxseed AOS, At AOS; Arabidopsis AOS

Figure 8

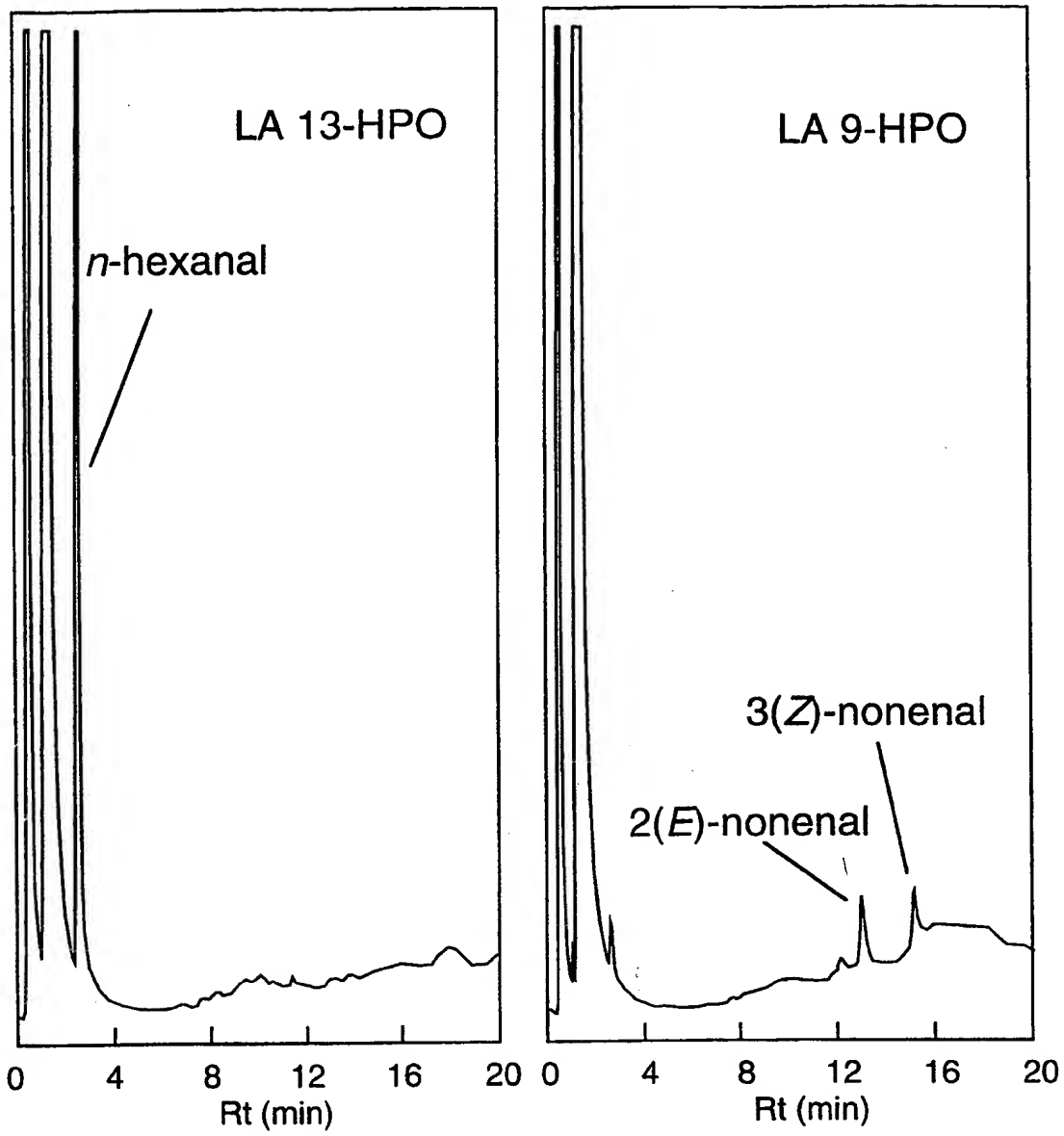

Figure 9A

Figure 9B

Figure 9

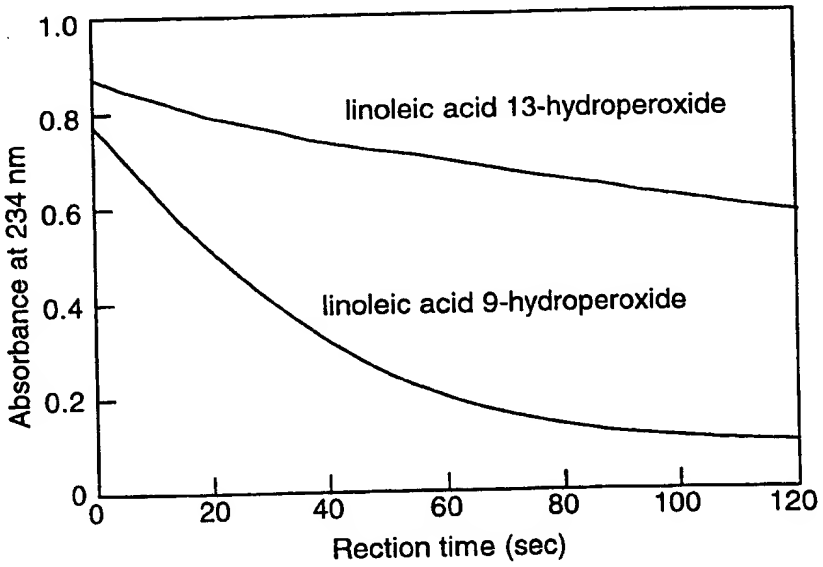
1.0
0.8
0.6
0.4
0.2
0
Absorbance at 234 nm
linoleic acid 13-hydroperoxide
linoleic acid 9-hydroperoxide
0
20
40
60
80
100
120
Rection time (sec)

Figure 10